ED 152 826	TM 007 006

TITLE Women Content in Units Force Development Test (MAX WAC).
INSTITUTION Army Research Inst. for the Behavioral and Social Sciences, Arlington, Va.
PUB DATE 3 Oct 77
NOTE 143p.

EDRS PRICE MF-$0.83 HC-$7.35 Plus Postage.
DESCRIPTORS *Armed Forces; Data Analysis; *Females; *Field Studies; Group Tests; Manpower Utilization; Measurement Techniques; Military Personnel; Participation; Performance Tests; *Program Effectiveness; *Quotas; *Task Performance; Test Construction
IDENTIFIERS Army; Women Content in Units Force Development Tests

ABSTRACT

The effect of varying the percentages of female soldiers assigned to representative types of Test, Operation, and Evaluation units on the unit's capability to perform its mission under field conditions was assessed. Empirical data were provided to test the null hypothesis that increasing the proportion of women in selected units would not impair unit performance. The basic concept was to test task performance in a total of 40 combat support and combat service support companies which were broken down into five different types of units: Medical, Maintenance, Military Police, Transportation and Signal. Within each unit type, eight companies were selected and designated as experimental, control, or calibration. Two experimental companies were tested twice, at varying fills of enlisted women (EW); the time between tests was six months. The control company was also tested twice with the EW fill stabilized for both tests. Five calibration companies were tested only once, with whatever percentage of women they contained. The comparisons of major interest involve companies that went from 0% to 15% EW and those that went from 15% to 35% EW. On the average, the former showed a slight decrease in performance scores while the latter showed a slight increase in performance scores. In neither case, however, were the changes statistically significant. (Author/MV)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED152826

Women Content in Units Force Development Test

(MAX WAC)

3 OCTOBER 1977

TM007 006

U. S. ARMY RESEARCH INSTITUTE
for the
BEHAVIORAL AND SOCIAL SCIENCES

CN000-040
1/16/77

FOREWORD

The work reported here was undertaken as part of the Army's long range effort to explore the future role of women in the Army.

The MAX-WAC research findings show that the number of women (up to as much as 35%) had no significant effect on the operational capability of specific Category II and III company size units as measured by Army Training and Evaluation Programs (ARTEPs). Ideally, this suggests that increases in women can be applied to tested or observed units (Signal, Maintenance, Military Police, Transportation, Medical Companies). There are 174 such units in the Army organized under the identical or similar TO&Es as the units tested. Extrapolation of test results to these units shows that we could accept up to 6,000 more enlisted women than provided in current assignment planning. However, this extrapolation assumes unit performance, as measured by 72 hour ARTEPs, to be the sole consideration in assignment. Other considerations which must be included in the Army's planning are the following:

a. Ability of women to perform for prolonged periods under field conditions,

b. Enlisted personnel management policies, and

c. Cost effectiveness comparisons.

The MAX-WAC study was extremely useful and provides some insight to the US Army in evaluating the role of women. The MAX-WAC test in itself does not provide an empirical basis to objectively establish an upper bound on the potential number of women in support roles.

WOMEN CONTENT IN UNITS

TABLE OF CONTENTS

PART	TITLE	PAGE
I	EXECUTIVE SUMMARY	I-1
II	INTRODUCTION AND RESEARCH METHODOLOGY	II-1
III	RESULTS AND DISCUSSION	III-1
IV	TEST DIRECTORATE TEAM OBSERVATIONS	
	A. ARI INTRODUCTION	IV-1
	B. QUALITATIVE ANALYSIS OF SUBJECTIVE EVALUATION OF WOMEN CONTENT IN UNITS (MAX WAC) FTD	IV-1
V	OTEA REVIEW AND EVALUATION	V-1

PART I

EXECUTIVE SUMMARY

WOMEN CONTENT IN UNITS

BACKGROUND: In late 1974, DCSPER recognized that the question of Women content in TOE Units would be an important future issue. In July 1975 BG Wroth (DAPE-PB at the time) addressed a letter to GEN Rogers (then CG FORSCOM) requesting support for a 'Test of Women Content in Units.' GEN Rogers agreed. DCSPER then tasked the US Army Research Institute for the Behavioral and Social Sciences (ARI) to develop such a test. When the resources required for the proposed test had been better defined, FORSCOM requested that the Test Schedule and Review Committee (TSARC) and the Operational Test and Evaluation Agency (OTEA) approve the test. ARI developed an Outline Test Plan (OTP) as required. In the ensuing coordination period prior to acceptance by TSARC of the OTP, discussions were held addressing the issue of how many FORSCOM units would be required for testing. OTEA proposed fewer units than ARI, and sophisticated statistics were argued at length. In the end OTEA and ARI were in agreement. The first tests began in October 1976.

PURPOSE: The purpose of this research was to assess the effects of varying the percentages of female soldiers assigned to representative types of category II and III TOE Units on the capability of a unit to perform its TOE Mission under field conditions. The objective as stated in the OTP was to provide empirical data to test the null hypothesis that specified increases in the proportion of women in selected units would not impair unit performance.

APPROACH: The basic concept was to test a total of 40 combat support and combat service support companies. These companies were broken down into eight companies each from five different types of units (Medical, Maintenance, Military Police, Transportation and Signal). Within each unit type the eight companies were designated as experimental, control, or calibration. Two experimental companies were to be tested twice, at varying fills of enlisted women (EW). The time between tests was to be six months. The control company was also to be tested twice with the EW fill stabilized for both tests. Five calibration companies were to be tested only once, with whatever percentage of women they contained. These companies established the range of scores one might expect, and some provided an opportunity for evaluators to gain experience before testing the experimental companies. The major statistical comparisons, however, were made between companies which were tested twice. The test design for the eight companies of each type unit appears as follows:

FILL LEVEL OF ENLISTED WOMEN FOR EACH TYPE OF UNIT

Test	Experimental		Control	Calibration	
Season	1 Co	1 Co	1 Co	2 Co's	3 Co's
Fall 1976	0%	15%	% as found	% as found	
Spring 1977	15%	35%	Same		% as found

ARI was directed to use a standard operational Army test in assessing company performance. The recently developed Army Training and Evaluation Program (ARTEP) was chosen as a vehicle for measuring company performance. The ARTEP, which is replacing Army Training Programs (ATPs) and Army Training Tests (ATTs), was chosen because it is "performance-oriented" rather than "procedure-oriented." The ARTEP is normally conducted over a three-day period, and thus the duration for each field evaluation was three days. A total of 55 ARTEPS were administered (10 experimental and five control companies were tested twice and 25 calibration companies were tested once.). In addition to the ARTEPS, ARI administered collateral questionnaires to 6,070 of 6,963 personnel to obtain additional data.

MAJOR FINDING:

- The comparisons of major interest involve companies that went from 0% to 15% EW and those that went from 15% to 35% EW. On the average, the former showed a slight decrease in performance scores while the latter showed a slight increase in performance scores. In neither case, however, were the changes statistically significant. Performance differences between the first and second ARTEP administration were small enough to be caused by chance. An effect due to the change in content of women was not established. (Note: The ARI interpretation is that women soldiers, up to the percent tested, do not impair unit performance during intensive 72-hour field exercises. It is predicted that a repetition of this Force Development Test (FDT) with more companies, improved instrumentation, and better controls of extraneous factors would yield essentially the same conclusion.

SUPPLEMENTARY FINDINGS:

- Leadership, training, morale and personnel turbulence were perceived by company officers and evaluators as having a greater effect on unit performance than the percent of EW in the company. Half of these officers perceived that the percent of women in a company contributed five percent or less to the total performance variation among companies.

- Over 80% of the officers, NCOs and enlisted personnel in the units tested indicated the ARTEP was either "excellent" or "OK" as a means of assessing the company's capabilities.

- Eighty-seven percent of the soldiers in the units responded to the collateral questionnaires.

- Less than 11% of the respondents thought that important jobs involved in accomplishing their wartime mission were omitted.

- Over 66% of the officers and NCOs indicated that the ARTEP included enough tasks to adequately measure gender-related differences in performance.

- Over 92% of the EW were in pay grades E1-E4 versus only 70% of the EM. Senior NCOs were primarily male; few female NCOs were represented in the test.

- EW in the test had more academic schooling than EM.

- In this sample, for both junior and senior enlisted, EW were less likely than their male peers to be married. Interestingly, among junior enlisted, EW report being divorced almost three times as often as their male peers.

- Approximately two thirds of the officers, NCOs and enlisted personnel reported their company performed "Outstanding/Very Well."

- A comparison of the evaluator scores and self-ratings from the first to the second ARTEP showed agreement in the direction of score change in thirteen out of fifteen cases.

- Male officers and enlisted men did not rate the performance of women as high as they rated the performance of men; e.g., 68% of the officers rated the performance of women as "Outstanding/Very Well," 79% of the same group so rated the performance of men. The EW, on the other hand, rated their performance slightly higher than that of males.

- Approximately 80% of EW and EM rated the performance of their group, squad or section as "Outstanding/Very Well."

- There is a need to give instruction to NCOs and officers on EW problems, so that appropriate leadership may be provided.

- EW are dissatisfied with their uniforms, and field hygiene is a problem.

CONCLUSION: The MAX WAC FDT was difficult to accomplish because of the many variables, e.g., leadership, post policies, personnel turbulence, weather. OTEA (in a Review and Evaluation of the MAX WAC Study forwarded to Director of Army Staff on 8 August 1977) has commented on the variability of performance on individual ARTEP tasks, due to these and other factors. It is the opinion of the ARI professional staff, based on all the data collected, that another test with tighter controls and an expanded test design would yield similar results, i.e., little or no relationship between unit performance (as measured by the ARTEP) and the number of EW in the unit, up to the percent here tested. The EW observed in the units were motivated and doing an excellent job. EW accomplished physically demanding tasks by utilizing leverage and a peer helper when required. EW appeared to do better in units where they were treated as equals and the leadership was supportive. Finally, it must be remembered that the FDT was conducted during a 72-hour period and that this is not long enough to determine how well EW will endure under extended field duty. ARI is addressing the issue of 'extended field duty' currently in another research effort entitled, Women in the Army - REFORGER 77.

It is recognized that the MAX WAC effort is one of many inputs contributing to policy determinations regarding the utilization of women.

PART II

INTRODUCTION AND RESEARCH METHODOLOGY

WOMEN CONTENT IN UNITS

1. INTRODUCTION

In 1967, Congress removed the 2% limit on the number of women who could be in the military services. At that time, there were approximately 10,000 enlisted women in the US Army representing less than 1% of enlisted strength. There was a gradual increase over the next five years so that, at the inception of the all-volunteer force, enlisted female strength had increased to about 13,000 (little less than 2% of the enlisted strength of a reduced force level). Over the next four years, however, female strength tripled, so that, by the end of fiscal 1976, there were almost 44,000 enlisted women (accounting for more than 6% of Army enlisted personnel). Concomitant with the rapid expansion in the number of women, all but a score of MOSs (those in the combat arms) were opened to women. Current Army goal is 50,400 enlisted women by the end of fiscal year 1979.

2. PURPOSE AND SCOPE

The rapid increase in the number of female soldiers, and the opening of enlisted opportunities in many MOSs formerly not available to them, raised a number of questions about the proper utilization of women in the Army. In April 1975, the Army developed policy limiting the percentages of women in non-combat units based on the type and normal location of the unit under emergency (wartime) conditions. These percentages range from 0% for units which normally operate forward of the brigade rear boundary to 10% for units operating between division and brigade rear, and to 15-30% for units between corps and division rear. Units which operate behind corps rear are allowed between 25-45%, and those not expected to leave CONUS during an emergency between 25% and 50%.

Limiting the percentage of women by type of unit, including 0% of women in the combat arms, places constraints on the number of women that the Army may access and still provide fair and equitable career progression for both male and female soldiers. There is, at the present time, considerable pressure for all the services to examine the feasibility of using more women in their branches. A recent study issued by the Office of the Assistant Secretary of Defense (Manpower, Research Affairs, and Logistics) entitled, "Use of Women in the Military" identified two main sources of such pressure. First, there is a growing movement within our society to provide equal economic opportunity for American women including their integration into the military. Second, the all-volunteer force is facing a significant decline in the potentially available youth population because of the lowered birth rates in the 50's and 60's.

Once ceilings had been placed on female enlisted strength in Category II and III TOE units (combat support and combat service support), Department of Army began planning to assess the adequacy of these quotas in relationship to the overall female strength ceiling to 50,400. In July 1975, BG Wroth, Director of Plans, Programs and Budget in the Office of the Deputy Chief of Staff for Personnel, requested the assistance of the Commander, US Army Forces Command (FORSCOM) in testing the ceilings under field operating conditions. After receiving a FORSCOM pledge of support in the form of providing units for testing, ODCSPER tasked the Army Research Institute to proceed to develop and conduct a test (Women Content in Units) in conjunction with FORSCOM.

The Army Research Institute began the lengthy process of planning for a comprehensive, large scale field experiment during the fall of 1975. In early discussions with various DA agencies and individuals in both DA and DOD, a concern was expressed that the results of such a test might eventually have to bear close scrutiny in a court of law. The General Counsel cautioned, for example, that testing of units should be done using a standard operational test such as an Army Training Test (ATT) rather than a specially designed test which might be attacked as biased, either for or against women. During the planning stage, ARI was directed, since the proposed project constituted a major commitment of Army resources, to submit the research design to the Operational Test and Evaluation Agency (OTEA) for review and the final plan to the Test Schedule and Review Committee (TSARC) for approval.

The original ARI research design called for three sets of annual ATTs to be given at the beginning, intermediate and end points of a two year period. Guidance from DCSPER to ARI necessitated the compression of the research effort into an eighteen month period beginning in May 1976. However, the need for TSARC approval and related requirements prevented starting the test until October 1976 and necessitated a quite different test design that could be accomplished with two sets of measurements obtained six months apart.

The Outline Test Plan (OTP) presented to the working group TSARC that preceded the General Officer's 1976 Spring TSARC meeting called for the use of 30 units, 6 of each kind, to be administered two ARTEPs 6 months apart. The day before the General Officer's TSARC a reduction from 30 to 10 twice-tested controlled-fill companies was negotiated among the OTEA, FORSCOM, and DCSPER TSARC representatives. This reduction was in essential accord with the recommendations by OTEA that a pilot study precede the more expensive (particularly regarding troop participation) twice-tested, 30-unit design proposed by ARI. An additional 40 companies were to be designated non-interference companies. These non-interference companies were to be made available to the evaluators to observe at whatever ARTEPs FORSCOM conducted for these companies, but ARI would have no control over scenarios, time of conducting ARTEPs, or even whether the ARTEPs would be conducted in garrison or in the field. By mid-June, correspondence outlining non-negotiable minimum requirements to provide a cost-effective data collection effort was sent to OTEA. FORSCOM's concurrence with these requirements launched the women content in units (MAX WAC) Force Development Test in mid-July.

The 18 June correspondence became a supplement to the OTP approved by the General Officer's TSARC; the two documents constituted the MAX WAX charter and were the sole basis for obtaining troop and other support from FORSCOM, TRADOC approval for using ARTEPs, and technical advisory service from the schools. The supplement was integrated into the OTP to create the 29 Sep 76 version of the MAX WAC OTP that was approved by the Fall 1976 TSARC.

3. RESEARCH DESCRIPTION

a. Test Design.

Formulation of a scientifically sound research design, given the parameters imposed by "real-world" conditions, resulted in a methodology of somewhat limited scope but responsive to the basic question posed in the tasking by DCSPER. ARI attempted to isolate the effect, if any, of different percentages of enlisted female soldiers on the performance of combat support and combat service support companies during a short-term (3-day) field exercise. It should be emphasized that, in accordance with the charter given ARI, attention was directed primarily on unit, not individual, performance. Women who participated in the test were required to be MOS qualified. Furthermore, it was required that they be assigned throughout the company. To test the major hypothesis of the project, it was necessary to determine whether the company could accomplish the myriad tasks which collectively make up its stated mission.

Forty FORSCOM Category II and III company-sized TOE units participated in the test. They were located at 19 posts in CONUS and Hawaii. The five types of units chosen for study were as follows: Medical Company (TOE 8-37H), Military Police Company (TOE 19-77H), Maintenance Company (TOE 29-207H), Signal Company (TOE 11-37H), and Transportation Light-Medium Truck Company (TOE 55-67H). The eight companies of each type were placed in one of three groups; the experimental group, a control group, or a calibration group. Assignment to groups was made by FORSCOM, who had to consider the problems involved in meeting the requirement, later in the test, to increase the percentages of enlisted women in the experimental group to as much as 35% of ALO-1 strength.

The core of the experimental design was a repeated measures (longitudinal) approach in which a company would act as its own control. Thus, the companies assigned to the experimental group were tested first at one level of female enlisted fill and about six months later at a different level of fill. To assess the effect of testing the same unit twice, the control group was to be tested during the first cycle of tests, the personnel stabilized as much as possible, and then tested again during the second cycle of tests. The remaining companies were tested once, about half during the first cycle of tests and the other half during the second cycle. This last group, referred to as the calibration group, served at least three purposes. Since there was no time, given the milestones provided to ARI, to pilot test the instruments and procedures that were to be used, by scheduling some of these calibration companies first, experience could be gained before the testing of the experimental and control companies began. Secondly, the range of scores, if not especially narrow, would allow statistical calibration of the scores obtained by the other two groups. Thirdly, since the percentage of women in companies varied, cross-company comparisons could be made between percentage of women in a company and ARTEP scores.

b. Test Instruments.

(1) ARTEP (Selected Tasks).

To assess company performance in the field, ARI was directed to use a standard operational Army test. The decision to use the newly developed ARTEPs was made for several reasons. ARTEPs are written by the Army schools and sent out for comment as coordinating drafts. Revisions are then made on the basis of comments received from the field and an updated version is published subject to revision as additional comments, based on users' experience utilizing the ARTEP for organizing and conducting field training exercises, are received. It turned out that for each of the TOE support companies identified for inclusion in the test, an ARTEP existed in at least coordinating draft form and that field comments had already been received. On the basis of assurance from the schools that any revisions of these drafts would be minor, it was decided to use the ARTEP in the form available. Several of the ARTEPs were considered operational. In any case, the superiority of the ARTEP to the older ATT favored its use for evaluating the companies on the field exercises.

As mentioned above, ARTEPs are produced by service schools under the guidance of TRADOC Reg 310-2. They are intended to replace the ATTs and associated ATPs, and to serve revised TRADOC objectives. Where the ATT was procedure-oriented, the ARTEP is performance oriented. Further, the doctrinal concept of the ARTEP is not as a test (evaluation measure), but as a diagnostic tool for the commander to identify training needs for all sections of the company or battalion. In essence, the ARTEP is based on an analysis of the unit's mission and lists the various tasks the company must perform in accomplishing that mission. Guidance is provided for constructing a 3-4 day field exercise scenario to assess the company's ability to perform its mission. The tasks are evaluated only in terms of being satisfactory or unsatisfactory. Special permission was required, therefore, from TRADOC to develop, for this one time only, a procedure for scoring the ARTEP results.

The goal was to extract from each ARTEP a sufficient number of tasks to keep the company active as well as to require them to demonstrate competence in accomplishing tasks deemed especially critical to the unit's mission. The scenario had to weave these critical tasks, along with others, into a 72-hour exercise that would constitute a realistic test of all sections of the company with a minimum of task simulation. It was, of course, accepted that the threat imposed by an enemy--ambushes, aggressor attacks on unit perimeter, casualties to be processed by medical companies, etc.--required simulation. The critical tasks selected for each of the types of companies were submitted to TRADOC and FORSCOM for approval. After some adjustments were made, eliminating some tasks and adding others, a final approved list of tasks was developed for each type of company. Each task was analyzed in terms of the components of the overall task, the sub-tasks that needed to be evaluated in order to assign an overall performance score. Most of these sub-tasks were provided by the ARTEP.

ARTEPs do not provide for differential scoring of tasks; this is in keeping with the TRADOC policy of using them as training diagnostic tools. ARI scientists felt that the pass/fail system was not sufficiently sensitive for the purposes of this test. Accordingly, a two-part scoring procedure was developed to provide more detailed assessments of company performance. Tasks and the sub-tasks were first rated on four separate factors. Table 1 lists these four factors and the definitions provided to the evaluators. It was felt that these four factors would focus attention on the performance of enlisted soldiers which was of primary interest in the test, since most of the women involved were in the lower (E1-E5) enlisted grades.

TABLE 1

PERFORMANCE EVALUATION FACTORS

FACTOR	SYMBOL	DEFINITIONS
TEAMWORK	Tw	Effective cooperation and coordination of effort between individuals working on a common task. (If test module or sub-task is performed by a single individual, teamwork is not assessed.)
NEED FOR SUPERVISION	NS	Each individual demonstrated appropriate skills, knowledge and abilities for task and requires only minimal level of supervision. Each individual carries full share of workload and demonstrates capability of working independently.
TIMELINESS	T1	Task or mission accomplishment within a suitable or allowable length of time.
QUALITY OF WORK	QW	Mission accomplishment is judged with respect to the accuracy, correctness and efficiency of action and the quality of the product. How well was the job done?

In rating tasks and sub-tasks, the evaluators were instructed to use a three-level rating scale as shown below:

Score	Basis of Rating
1	Unsatisfactory
2	Satisfactory - Average to slightly above average
3	Outstanding

An example of a score sheet used by the evaluators for the MP companies is shown in Table 2. The critical task (called the Test Module here)

EVALUATOR SCORE SHEET

Page: 3

ARTEP 19-77 MP CO (TOE 19-77H)

Date:

UNIT:

MISSION: Control Traffic at River Crossing

TEST MODULE & REF. PAGE: Control Traffic (Crossing Area), F-1-3

SITUATIONAL DESCRIPTION	TRAINING/EVALUATION SUB ITEMS	Tw	NS	T1	QW	Score	COMMENTS
Each platoon leader completes his/her coordination and reconnaissance, issues his/her order and positions his/her personnel in accordance with the time frame established by the evaluator. Each bridge site is subjected to sniper fire and intermittent artillery fire. A few personnel in each MP platoon are wounded and two platoon vehicles are damaged by enemy fire during the course of the river crossing.	MP personnel assigned duty in the crossing area fulfill the following: 1. Operate dispersal areas where vehicles can be halted and dispersed.						
	2. Dispatch vehicles from the dispersal area to the crossing site in accordance with the capability of the crossing areas.						
	3. Personnel are on position in time, strength and location to support river crossing.						
	4. Personnel are fully conversant with mission and duties to which assigned.						
	5. First aid and medical evacuation is provided for wounded personnel.						
	6. Timely reporting and replacement of equipment damaged by enemy fire is effected.						

TABLE 2

is keyed to the ARTEP (ARTEP 19-77, Test Edition, dated March 1975) covering this type of Military Police TOE. Evaluators were instructed to consider the sub-tasks first, rating each on the four factors (by assigning either a 1, 2 or 3) before giving each sub-task an overall score in the box at the far right. Having rated all sub-tasks, they were then required to rate the critical task, e.g., "Control Traffic (Crossing Area), F-1-3," on the four factors separately before assigning an overall score for that task (the large square directly above "score").

(2) Collateral Research Measures.

ARI did not have an opportunity, within the time frame specified for conducting the test, to pilot test instruments and procedures. As an aid to interpreting the test results, a set of questionnaires was developed to collect additional information, attitudes and opinions from the participants. These questionnaires were designed to provide insights into organizational and individual factors that impact on the effect that content of women has on morale and performance in these combat support and combat service support units. There were four different questionnaires:

(a) Field Questionnaire. A short questionnaire was administered to all enlisted personnel towards the end of the exercise while they were still in the field. It required 10 to 15 minutes to complete and was designed to elicit opinions about the ARTEP and about how well the company performed.

(b) General Enlisted Questionnaire. Usually at the beginning of the week following the exercise, all enlisted company personnel were administered a more comprehensive questionnaire. This instrument repeated the field questionnaire first, to assess any changes in opinions after getting back to garrison and having a chance to clean up and catch up on sleep. In addition to obtaining some personal history (demographic information) from the respondents, the questionnaire addressed a variety of issues. These included attitudes towards women and the role of women, confidence in male/female peers, opinions on the impact of women on unit effectiveness, and personal views on combat. Information was sought about MOS mismatch, views about deployability and tasks requiring strength and stamina. The questionnaire required one to one-and-a-half hours to complete.

(c) Supervisor's Questionnaire. Certain selected first-line supervisory NCOs were given a separate questionnaire tailored to their position in the company. It was designed to explore duty assignment practices with special attention to whether gender influences their organization of work crews. It took about an hour to complete.

(d) Officer's Questionnaire. Beginning with the Spring test cycle, a questionnaire was given to the company's officers. An attempt was made to obtain completed questionnaires from the officers involved in the already completed ARTEPs by mailing them copies to be completed and returned to ARI. The content of this questionnaire was similar to the general enlisted questionnaire with additional questions about command practices.

The general enlisted and the officer questionnaires address two issues of some importance in light of some of the limitations and problems of the test: the validity of the ARTEP, and peer and leadership opinions of the performance of women. ARI was concerned about the participants' perception of the ARTEP as a measure of unit capability to perform its wartime mission, especially since, in some cases, ARTEPs were being used for the first time in the field. Short of sending a unit into combat after being evaluated on an ARTEP, valuable estimates of validity may be obtained from participants' observations. The collateral instruments also provided the opportunity, in a general way, to multiply the evaluators' judgments many times by getting opinions about women's performance from both peers and leaders. The judgments provide independent secondary criteria about the performance of enlisted women in the field. It is possible to relate these judgments to a number of other variables measured during the test.

(3) Management Information. At the conclusion of the first test cycle, with the experience gained in conducting more than 20 field exercises, the Directorate decided to systematically collect additional survey type data which would be of general interest in the management of female soldiers. Questions were added to the enlisted questionnaire addressing the issues of sole parenthood, deployability, pregnancy and hygiene problems in the field, physical strength requirements found taxing for women, and continuity of supervision when moving from garrison to the field. Each of these issues was perceived as a common problem area in the utilization of women in the Army which had not been specifically addressed in the original questionnaire.

c. Training Package. A major concern, for the companies undergoing repeated testing with same scenario was the effect feedback from the first administration might have on the second test. It was felt that poor performance on tasks during the first test could cause the conscientious company commander to concentrate training time and resources to correct the deficiency before the second test. Two measures were taken to attempt to counter this possibility. In the first place, the design plan called for all twice-tested units to be given a 60-day training period prior to each ARTEP. The required female level of fill was to be attained before the start of the 60-day period. A training package was delivered to the company before the beginning of the training period; the package contained a detailed Letter of Instruction (LOI), the school-produced ARTEP and the summary of the scenario to be used on the field exercise. Additionally, arrangements were made for all reference material listed in the ARTEP (FMs, TMs, TCs, etc.) to be delivered to the company by pin-point distribution.

The training package and training lead time were provided to allow companies, theoretically at least, enough time to prepare adequately for the first ARTEP. A summary of the scenario was given the company commander under the philosophy of "no secrets" on the field exercise so that the test would remain an open test of how well enlisted soldiers know their jobs (and not how well leaders react to unexpected situations).

It was felt that given that amount of open information, there would be less chance for a company to do so poorly on the first ARTEP that remedial training would have a significant effect on the scores obtained on the second test. The second measure taken to ameliorate a "training" effect from the first to second administration was to require the company, during the first training cycle, to maintain a training log and record the actual amount and kind of training conducted. The log was handed over to the evaluation teams at the conclusion of the first ARTEP. Prior to the beginning of the second training cycle, the log was returned and the companies instructed not to exceed the time or kind of training given during the first training period.

The five companies of each type tested once (calibration group) were given the same amount of time to prepare for the ARTEP and the same materials and information (Training Package). They were also required to maintain a training log in order to create comparable test conditions for all companies.

d. Test Directorate.

A Test Directorate was established, with a Test Director (COL) and a Deputy Test Director (LTC), consisting of five Evaluator teams (called Umpires in the OTP). Each team was to be headed by a branch qualified Team Chief, in all cases but one a Major, with command experience in that branch. The remainder of the team consisted of one branch qualified CPT, one combat arms CPT and one female CPT, branch immaterial. An administrative NCO (E8) and several civilian clerk typists completed the Directorate personnel. During the Fall test cycle, they were stationed TDY at ARI headquarters in the Washington, D.C. area. After the first of the year, about half of them returned to their home stations, while the other half remained in Washington. Those who had returned to their home station went TDY to each ARTEP location and periodically to ARI for conferences and to deliver completed instruments.

Coordination of ARTEPs was effected by the Directorate, first through personal visits by Directorate members and later by telephone and messages. A personal visit was made at least once before the ARTEP to every unit involved in the test. Direct communication was authorized by FORSCOM between the Directorate and all levels of installation command.

Conduct of each ARTEP was under the direction of a local post evaluation team who were required to use the ARI-developed scenario. The Directorate evaluation teams were instructed to remain as unobtrusive as possible while still ensuring that the scenario was adhered to as strictly as possible. The local evaluation teams were not informed of the evaluations made by the Directorate teams nor were they asked to provide the Directorate with their evaluations. This was in keeping with the promise of confidentiality of data made during initial coordination visits. In general, cooperation between local evaluators and Directorate teams was excellent, as was installation support. It should be noted that for the first test given the twice-tested units, and for all the once-tested units, the ARTEP constituted an official evaluation.

conditions. Whenever possible, the same members of each four-officer team observed and scored the same critical tasks. A promotion and transfer, a married pregnancy, a resignation and a retirement forced the change of several evaluators. It was felt, however, that this unpreventable personnel turbulence did not seriously affect consistency of the evaluations. Another potentially serious problem concerned the lack of time for the evaluators to gain experience through pilot testing and fix their own evaluation standards. To counter the possible tendency for personal judgment standards to "drift" as more experience was gained in the field, the evaluators were instructed to try to adhere to their first standards. If their initial scores appeared to be too high or too low in the light of later experience, evaluators were told to continue to use those early standards. After each ARTEP, the Director and usually the Deputy Director conducted a lengthy debriefing, partly to reinforce the need for consistency over the entire course of testing.

e. Scenario Development. This test focused major attention on the contribution made by the job performance of enlisted men and women, especially in the first four grades, to overall unit performance on the ARTEP. Therefore, scenarios were written for the five types of companies to highlight the work of these soldiers. The scenarios were written with three major considerations in mind. (1) Each was written in accordance with a SCORES mid-intensity European scenario. (2) Each was written to reduce the decision-making role of the company leadership. This was done to try to standardize the test procedures across all eleven ARTEPs (within each type of unit; e.g., Med, Trans), to provide a context meaningful to decision-makers, and to focus performance measurement on the grade levels in which women soldiers were already present or could be introduced. The ARTEP had to be administered under conditions that permit meaningful comparisons of ARTEP scores across companies of the same type. (3) Each scenario had to contain many tasks in addition to the critical tasks rated by the evaluators, in order to ensure that the whole company was kept occupied during the entire 72 hours. Although soldiers were not stressed or taxed to the limit, a realistic test required that there be little nonproductive time. In line with this philosophy, only genuinely malfunctioning equipment was to be repaired or actual messages transmitted. Simulation was used only when it was impractical to have the real thing.

4. TEST CONDITIONS

a. Schedules.

Testing began in fall 1976 and the second cycle of tests followed approximately six months later in Spring 1977. There were two companies within each type of unit in the experimental group. One company was tested first at 0% EW and about six months later at 15%. The other experimental company of the same type was first tested at 15% EW and then at 35% EW. The control company was tested in the fall with existing percentage of EW. The personnel in the company were then stabilized, as much as possible, and tested again in the spring. The five companies in the calibration group were tested with existing percentage of EW, two of them in the fall and three in the spring. The eight companies of each type, then, were distributed into the three groups as described above. The basic design is presented, for any single company type, in Table 3 below:

TABLE 3

FILL LEVEL OF ENLISTED WOMEN

Test Season	Group 1 EXPERIMENTAL		Group 2 CONTROL	Group 3 CALIBRATION	
	1 Co.	1 Co.	1 Co.	2 Co's.	3 Co's.
Fall 1976	0%	15%	% as found	% as found	
Spring 1977	15%	35%	same		% as found

Fifteen of the companies (three of each type) were tested twice, while 25 (five of each type) were tested once for a total of 55 field tests. Ideally, the twice-tested units would have had the specified six-month interval between tests. In reality, schedules had to conform with various installation requirements and there was some variability in the intervals between the two tests. Testing began in early October 1976 and concluded in late June 1977. The schedule of tests is presented in Table 4.

TEST SCHEDULES

DATE	SIGNAL	TRANS	MEDICAL	MAINT.	MIL. POL.
4- 8 OCT		CONTROL		CONTROL	
11-15 OCT			CALIB.		CALIB.
18-22 OCT		CALIB.			CONTROL
25-29 OCT	CONTROL			CALIB.	
1- 5 NOV				EXP.15%	CALIB.
8-12 NOV		EXP.0%	EXP.15%		
15-19 NOV	EXP.0%	CALIB.	EXP.0%	CALIB.	EXP.0%
22-26 NOV					
29 NOV-3 DEC	CALIB.		CALIB.		
6-10 DEC				EXP.0%	
13-17 DEC	CALIB.				
24-28 DEC					EXP.15%
31 JAN-4 FEB			CONTROL		
7-11 FEB					
14-18 FEB	EXP.15%	EXP.15%			
21-25 FEB					
28 FEB-4 MAR					CALIB.
7-11 MAR				CALIB.	
14-18 MAR		CALIB.			
21-25 MAR					CALIB.
28 MAR-1 APR					
4- 8 MAR		CALIB.	CALIB.		
11-15 APR	CALIB.				CALIB.
18-22 APR		CALIB.	EXP.15%		
25-29 APR	CALIB.		CONTROL	CONTROL	CONTROL
2- 6 MAY	CONTROL	CONTROL		CALIB.	
9-13 MAY					EXP.15%
16-20 MAY	CALIB.		CALIB.		
23-27 MAY		EXP.15%	CALIB.	EXP.35%	EXP.35%
30 MAY-3 JUN				CALIB.	
6-10 JUN	EXP.35%		EXP.35%		
13-17 JUN	EXP.15%	EXP.35%			
20-24 JUN				EXP.15%	

Unit designations and installation identifications are omitted to ensure confidentiality of the results. The only "official" evaluation of these units was made by local evaluators who actually conducted the field exercises. Their evaluations were made separately and, in accordance with the spirit of TRADOC doctrine regarding ARTEPs, were provided to unit commanders as diagnostic feedback telling them in which areas they needed to concentrate their training time. The scores awarded by the Test Directorate teams conducting the test were intended for research purposes only and were not divulged outside of ARI. A pledge of confidentiality of research data was considered fundamental to successful conduct of the field experiment.

b. Assignment of Women.

The Outline Test Plan defines the conditions governing the assignment of women in those units in which the level of fill was controlled. The most important consideration was that females be assigned in a large number of MOSs contained in each company's TOE; otherwise, the entire purpose of the test would be invalidated. MILPERCEN was given the responsibility per HQDA Ltr, 9 Nov 76, for assigning only MOS-qualified women to slots designated as interchangeable by the TRADOC study. To provide guidance for MILPERCEN, ARI analyzed the MOS distribution for each TOE, grouped MOSs together, and specified the number of positions to be selected from each group to meet the 15% and 35% fill levels. Table 5 reproduces this guidance for each type of company. It should be noted that, of the 49 MOSs in the selection list shown in Table 5, women actually served in 43 of them.

A second requirement specified in the OTP was that "all personnel available for duty at the time of the ARTEP shall participate in a manner appropriate to his/her MOS." The OTP directed that commanders not allow their companies to leave women behind in the company area during the ARTEP, "to handle essential administrative or urgent installation support--except for such reasons as illness or physical injuries." To ensure that the companies "don't leave the women behind," they were required to supply unit rosters and to account for all company personnel. The stated goal was to have twice-tested units (experimental and control groups) filled to within 90% of ALO strength and the once-tested units within 80%. On the average, the actual percentages of authorized personnel in the field was 87.4% for the twice-tested units and 86.8% for the once-tested units. The range for the former was from 58% to 106% while for the latter the range was from 62% to 116%. Although the number of personnel in the field did not always meet the requirement specified for the test, the number of enlisted women as a percentage of those in the field was within acceptable limits. In the presentation of results later in this report, data will usually be plotted against percentages of women out in the field derived from the following formula:

$$\% \text{ of women} = \frac{EW}{EW + EM} \times 100$$

c. Control of Variables.

A field experiment of this magnitude involves so many variables which might impinge on the dependent measures (i.e. unit performance) that control of all variables is extremely difficult, if not impossible. In the absence of direct control and of pilot work, one recourse is to measure (or record) as many aspects as possible of the conditions under which the tests are conducted and attempt to effect statistical control of these variables. A thorough discussion of the problems connected with the test is found in the Test Design Plan (TDP). The TDP also outlines the rationale and approach to the major statistical analyses

TABLE 5

MOS QUOTAS FOR SELECTION OF FEMALE SOLDIERS

TYPE UNIT & TOE	MOS		QUANTITY	
			15%	35%
MAINTENANCE (29-207H)	31E, 31J, 36G, 36K		1	3
	45L		1	2
	41C, 51A, 62F, 62M		1	2
	51L, 63G		2	4
	62B		4	8
	45K, 63C		1	3
	44B, 44E, 45B		1	3
	63B		2	5
	63J		1	3
	52B, 52D		3	7
	63F, 64C		1	3
	63H		7	17
	94B, 71B, 75B, 76P, 76V, 76Y		3	8
	76D		3	6
	ALO Strength = 212	Total	31	74
MEDICAL (8-37H)	52B, 63B		1	1
	75B, 76[illegible] 76Y, 94B		2	4
	91C, 91D, 91E, 91G, 91P, 91Q, 92B		3	6
	91B		5	12
	ALO Strength = 72	Total	11	25
MIL. POL. (19-77H)	31B, 36K, 52B, 63D		2	6
	71B, 75B, 76D		1	3
	76Y, 94B		1	3
	95B		21	49
	ALO Strength = 173	Total	26	61
SIGNAL (11-37H)	05C		1	2
	05E		2	7
	05F		2	4
	31M		10	24
	72C		3	6
	72E		8	19
	75B, 76D, 76Y		1	2
	94B		2	4
	ALO Strength = 193	Total	29	68
TRANSPORTATION (55-67H)	36K, 52B, 63B, 63F		6	14
	64C		8	19
	75B, 76D, 76Y, 94B		3	7
	ALO Strength = 117	Total	17	40

for testing the hypotheses posed in the design of the test. Ultimately, the object of the research design for the test was to eliminate or isolate (that is, identify and measure) all those factors which might affect unit performance on the ARTEP except the variable of interest, the percentage of enlisted women.

5. TEST OBJECTIVES AND PLANNING CONSIDERATIONS

The tasking order to ARI from DCSPER stated that, ". . . it is planned to fill selected CAT II and III units with the recommended maximum percentages of female enlisted soldiers in order to test unit performance under field operational conditions." Thus, the original task, as stated by DCSPER, was to field test the quotas promulgated earlier by DA limiting the percentage of women in CAT II and III TOE units. In arriving at a suitable research design, a number of factors were considered. These are briefly discussed below:

a. The need to be able to generalize results.

At the time the test was being planned, women were being (or had been) trained in a wide variety of MOSs and assigned in many Category II and III companies. A sufficient number of women had had training in newly available MOS skills to make testing of several kinds of support companies feasible. In light of the task given ARI, it was necessary to include as many different support companies as possible to be able to generalize the results to the maximum extent.

b. The need for comprehensive inclusion of MOSs.

If women were used only in traditional MOSs or kept back in garrison, the whole point of the test would be missed. DA policy and doctrine permits women to train in all but the combat arms MOSs while, at the same time, limits the percentage of women in combat support and combat service support MOSs. Expansion of the number of women in the Army would increase the number entering non-traditional jobs so that any test of the utilization of women would have to include women working in these jobs.

c. The need for standardized testing.

Guidance to ARI during initial discussions included using an "off-the-shelf" operational test, such as an ATT, to avoid later charges of bias if a specially constructed test were used. The ATTs, however, varied widely with respect to the amount of detail provided, in the amount of scoring possible beyond a pass/fail judgment, and in the repeatability of prescribed tests. Additionally, during the planning stage, ATTs were being replaced with ARTEPs (Army Training and Evaluation Program). ARTEPs, because they are performance-oriented rather than procedure-oriented, were desirable vehicles for conducting the tests. However, not all of the ARTEPs had been issued or field validated.

d. The limitation on resources.

A field experiment of this magnitude involves a host of variables capable of affecting the major dependent variable or measure and difficult to control. Statistical confidence can be increased by increasing the number of units tested, but costs and possible disruption of mission accomplishment place constraints on the number of units that can be realistically involved in the test.

e. The need to control the number of women in units.

By the research design, the independent variable was the proportion of enlisted female soldiers in the company. It became necessary, therefore, to structure test companies with specified levels of qualified women soldiers and sometimes change the level of fill. Since ARI was charged with measuring the impact of female soldiers on unit performance, women had to be assigned across the entire list of enlisted duty positions and not concentrated in traditional jobs. In this way, women would be in a position to affect performance throughout the company and not in limited activities of the company.

f. The need for expert evaluation.

It was determined that evaluation of company performance under field conditions required expertise resident only in military personnel. This recognition dictated formation of teams of active duty military personnel who could be stabilized throughout the course of the test. Continuity of the evaluation teams and careful selection of team members was a major concern during the design phase.

g. The lack of female NCOs and officers.

At the time the research design was being developed, it was recognized that there were simply too few women in leadership positions, both commissioned and enlisted, to include this factor in the design. With the time available for the test, it did not appear possible to either manipulate or control unit content of women in leadership positions.

h. The need for a reliable measuring device.

A major concern was the need for a scoring system for measuring unit performance that could differentiate between levels of performance and that could be defended on psychometric grounds. The pass/fail procedure of both ATTs and ARTEPs was not deemed adequate to produce data which would assure that obtained differences were large enough to have practical significance and would have statistical significance as well. It was recognized that the time constraints placed on the test would not provide enough time to pilot test and subsequently adjust and fine tune the scoring procedures.

PART III

RESULTS AND DISCUSSION

WOMEN CONTENT IN UNITS

1. RESULTS

a. Introduction.

Unit performance, as measured by performance of selected tasks during a three-day ARTEP, constituted the principal dependent variable in this field experiment. The scores awarded by evaluators to the various critical tasks formed the basis for arriving at a measure of company performance. For purposes of analysis, equal weight was given to each of the rated tasks and a simple arithmetic average was used to represent each company's score. Some data was missing where tasks were not scored for a variety of legitimate reasons (such as non-availability of equipment to repair). Although there are a number of statistical techniques for handling missing data, simple averages have been used. Statistical comparisons of scores adjusted for missing data by more complex techniques would change the findings only an insignificant amount.

b. ARTEP Validity.

Several analyses of collateral research data will be discussed before presenting the data based on evaluator scores. The newness of ARTEPs, the inability to pilot test the procedures, and the short preparation time prompted the inclusion of a number of questions in the collateral research instruments which asked opinions about the ARTEP as a vehicle for assessing a company's ability to accomplish its mission. It is instructive to consider the opinions of those involved as a measure of the face validity of the exercise, in the absence of more traditional measures of test validity. It is obvious that the opinions of some participants, such as those with greater experience including combat or wartime service, lend greater credence than the opinions of those with little military experience. Before considering the major findings on performance, therefore, data bearing on support for the validity of the ARTEP as a measure of unit proficiency will be presented in some detail, with some background information about the respondents.

Officers and enlisted service members were asked what they thought of the ARTEP as a means of assessing a company's ability to perform its wartime mission. It was recognized that only a small proportion of the company personnel would have experienced wartime conditions; i.e., Viet Nam, and some caution would have to be used in interpreting results.

Thus, the data from officers (especially 03) and senior enlisted (E5-E9) are more likely to reflect wartime experience than the data from more junior officers and lower ranking enlisted soldiers. Table 6 presents the results from this question, by rank, with enlisted personnel further divided into lower and higher rank. The five response categories have been collapsed for ease of presentation. As can be seen, over 80% of the respondents thought the ARTEP was either "excellent" or "OK" as a means of assessing the company's capabilities.

TABLE 6

IS THE ARTEP A GOOD MEASURE OF WARTIME PERFORMANCE? (in %)

Response Alternatives	OFFICERS (N=138)	E5-E9 (N=1603)	E1-E4 (N=4320)
Excellent/Pretty Good	55.8	53.1	45.1
OK	37.7	34.9	42.0
Not Very or Any Good	6.5	12.0	12.9

The second question asked the respondents whether the ARTEP (the scenario derived from the ARTEP and driving the exercise) covered most of the important tasks the company has to be able to do in a wartime situation. The results are presented in Table 7 with the response categories collapsed again for ease of presentation. Few of the respondents thought that important jobs involved in accomplishing their wartime mission were omitted.

TABLE 7

DOES ARTEP COVER EVERYTHING IMPORTANT? (in %)

Response Alternatives	OFFICERS (N=137)	E5-E9 (N=1617)	E1-E4 (N=4301)
Everything or About Everything Important	62.8	68.4	67.1
Most of the Important Things	30.7	21.5	23.0
Few or Any of the Important Things	6.5	10.1	9.9

Since the general purpose of the test was well known by most participants, the next question asked them if the ARTEP included enough tasks that would show gender related differences in performance. Table 8 presents the results from this question. Although about two thirds of each rank category thought enough tasks were included, almost one-third thought that there were not enough of these tasks. Two open-ended questions followed asking which tasks should have been included and which left out. These data have not been content analyzed but will be covered in a later ARI Technical Report.

TABLE 8

ENOUGH GENDER SENSITIVE TASKS ON ARTEP?
(in %)

Response Alternatives	OFFICERS (N=131)	E5-E9 (N=1906)	E1-E4 (N=3891)
Too Many Tasks	1.5	4.7	6.8
About Right Number	67.9	66.5	64.8
Not Enough Tasks	30.5	28.8	28.4

The data reviewed above offer some assurance that participants thought that the ARTEP-based field exercise constituted a generally valid measure of the company's ability to perform its TOE mission. The conclusion is that, although it is not perfect, the ARTEP is the product of expert judgment and is perceived by soldiers, both commissioned and enlisted, as valid. The positive endorsement of soldiers and leaders actually involved in the 55 field exercises lends credibility to the use of the ARTEP-based scenario as a measure of unit performance. The lack of complete unanimity of opinion suggests that improvements can be made, but, given the newness of the ARTEPs, the positive nature of the responses to these questions suggests that it is unlikely that gross errors would be made using the ARTEP as a basis for measuring unit performance.

c. Sample Characteristics.

The collateral research questionnaires were a source of information about the people involved in the test. Although those given the questionnaires were informed that they did not have to fill them out, most complied, and missing data tended to be unsystematic; i.e., a few questions per questionnaire were not answered. Accordingly, the data that follow are self-reported and obtained from anonymous questionnaires.

(1) Enlisted Background Characteristics.

The first background variable examined was the distribution of paygrades of enlisted soldiers. Table 9 presents the data for paygrade by gender. As might be expected, given the types of companies tested and the recent entry of women into many of the MOSs in these companies, over 92% of the women were in paygrades E1-E4 versus only 70% of the men. Senior NCOs were primarily male and very few female NCOs were represented in the test.

TABLE 9

PAYGRADES OF ENLISTED SOLDIERS

	MALES		FEMALES	
Paygrade	N	%	N	%
E1	139	2.7	22	2.6
E2	748	14.4	188	22.5
E3	916	17.6	231	27.7
E4	1822	35.0	330	39.5
Subtotal=	3625	69.7	771	92.3
E5	931	17.9	54	6.5
E6	404	7.8	7	.8
E7	193	3.7	3	.4
E8	40	.8	0	0.0
E9	9	.2	0	0.0
Subtotal=	1577	30.3	64	7.7
Total=	5202		835	

The second variable examined was the age of the enlisted soldiers. Table 10 presents the age data broken down separately by gender and enlisted level; i.e., E1-E4 and E5-E9. The female soldier in the lower enlisted paygrades is comparable to the male in age, even though the minimum enlistment age is higher for women than for men. The average reported age of male soldiers E1-E4 was 21.09 whereas the average for women was 21.41. Reflecting the longer service of males is the fact that the average age of male NCOs (E5-E9) was 28.87 while for females (E5-E7) it was 24.29 years.

Two questions examined the educational background of enlisted soldiers. The first question simply asked for the number of years of schooling the respondent had. The results from this question are presented in Table 11, again broken down for the two levels of enlisted ranks and gender.

TABLE 10

AGE OF ENLISTED SOLDIERS
(in %)

	E1 - E4		E5 - E9	
Age	Males (N=3453)	Females (N=748)	Males (N=1521)	Females (N=63)
17-18	10.5	7.5	.3	1.6
19-20	38.7	44.0	1.4	11.1
21-22	28.2	22.9	11.0	28.6
23-24	12.0	11.8	18.3	20.6
25-26	5.4	6.7	14.0	19.0
27-28	2.5	3.2	11.2	7.9
29-30	1.2	1.3	8.9	4.8
31-35	1.1	2.5	17.1	0.0
36-40	.2	0.0	13.4	6.3
41-45	.1	0.0	3.6	0.0
46-50	.05	.1	.9	0.0

TABLE 11

YEARS OF EDUCATION OF ENLISTED SOLDIERS
(in %)

	E1-E4		E5-E9	
Years Education	Males (N=3529)	Females (N=759)	Males (N=1557)	Females (N=59)
Less than 10	3.4	.3	.9	1.7
10	5.0	1.4	1.3	0.0
11	7.6	1.3	2.2	0.0
Subtotal=	16.0	3.0	4.4	1.7
12	59.9	65.0	63.8	61.0
13	12.9	15.2	16.4	18.6
14	7.5	11.7	10.9	10.2
15	1.5	2.8	2.1	1.7
16	1.6	1.6	1.9	5.1
17	.3	.1	.2	1.7
18	.1	.5	.3	0.0
19	.03	.1	.1	0.0
Mean # Yrs.	12.12	12.53	12.47	12.68

At the E1-E4 enlisted level, the largest difference is for those reporting less than 12 years of schooling where 16% of the males but only 3% of the females report less than 12 years. This difference proved significant by chi-square test* ($X^2 = 88.94$, $p<.001$). Additionally, the women in the E1-E4 group report post-high school attendance more often than males, 32% vs 24% ($X^2 = 21.16$, $p<.001$). These differences are less pronounced among senior male and female enlisted. On the whole, however, the females in the sample had more schooling than the males. The difference in educational attainment is highlighted in Table 12 which presents gender and rank for the highest diploma or degree attained. At both enlisted levels, E1-E4 and E5-E9, women have had more conventional education than their peers. At the E1-E4 level, X^2 comparison (males vs females) of H.S. Graduate or beyond with No High School and GED yields $X^2 = 73.88$ ($p<.01$). A similar comparison at the E5-E9 level yields $X^2 = 7.99$ ($p<.01$).

TABLE 12

HIGHEST EDUCATIONAL LEVEL ATTAINED
(in %)

	E1-E4		E5-E9	
Educational Level	Males (N=3537)	Females (N=762)	Males (N=1551)	Females (N=64)
No High School	13.4	0.5	2.5	0.0
GED	12.0	10.5	22.3	9.4
H.S. Graduate	67.3	79.3	64.4	79.7
Assoc. Degree	4.8	6.4	7.4	4.7
Bachelor Deg.	1.5	1.3	1.7	6.3
Grad. Degree	0.9	2.0	1.7	0.0

The marital status of the respondents is presented in Table 13 by rank and gender. For both junior and senior enlisted, females are less likely than their peers to be married (for junior enlisted, $X^2 = 45.05$, $p<.001$; for senior enlisted $X^2 = 43.3$, $p<.001$). Interestingly, among junior enlisted, women report being divorced almost three times as often as their male peers ($X^2 = 56.68$, $p<.001$). The typical male NCO is seen as married, whereas less than half of the female NCOs are or have been married. Caution should be exercised in interpreting these data, however, due to the small number of women in the senior enlisted ranks in this sample.

---

* Chi-square is a computed statistical value obtained from a data table which, with the associated degrees of freedom, can be checked against published tables to determine if a relationship exists which can be declared to be greater than could be expected by chance at the indicated level of confidence. At the $p<.001$ level of confidence, the possibility of the results occurring by chance (when no relationship really exists in the parent population from the sample was drawn) is less than one in a thousand.

TABLE 13

MARITAL STATUS OF ENLISTED SOLDIERS
(in %)

	E1-E4		E5-E9	
Marital Status	Males (N=3517)	Females (N=760)	Males (N=1542)	Females (N=63)
Married	40.7	27.6	79.5	44.4
Separated	2.6	2.6	3.3	1.6
Never Married	53.1	60.4	10.8	52.4
Divorced	2.8	8.6	6.0	1.6
Widowed	.8	.8	.5	0.0

A great deal of interest and concern has been expressed recently about the ability of female soldiers to meet the physical requirements of the jobs they are being trained to do. As a part of the collateral research effort, respondents were asked their height and weight along with some questions about the physical demands of their jobs. The latter data have not been analyzed to date, but the height and weight data are presented in Table 14.

TABLE 14

HEIGHT AND WEIGHT OF ENLISTED SOLDIERS

	FEMALES		MALES	
	E1-E4	E5-E9	E1-E4	E5-E9
HEIGHT (in.)				
Mean *	65.26	65.63	70.34	70.35
Median *	65.10	65.25	70.50	70.53
Mode *	64.00	61.00	71.00	71.00
WEIGHT (lb.)				
Mean	132.81	135.29	165.86	176.50
Median	130.26	132.00	164.57	174.90
Mode	130.00	110.00	160.00	160.00

* The mean, median and mode are each measures of central tendency. The mean, or arithmetic average, is the sum of all measures divided by the number of measures. The median is the numerical value exceeded by one half of the measures, and the mode is the single numerical value which has the highest incidence of occurrence.

(2) Officer Background Characteristics.

Beginning with the second cycle of testing, a questionnaire was constructed to be given to the officers of each company. The officers who had been involved in the first ARTEP were picked up on the second test if they were still with the company. Those officers tested only during the first cycle (that is, with a calibration group company) were mailed questionnaires. Approximately 75% of the officers in the tests completed questionnaires.

Some personal background information was requested from the officers. Table 15 summarizes the data obtained from 139 questionnaires.

TABLE 15

OFFICER QUESTIONNAIRE DATA

Rank	2LT	1LT	CPT	Missing	Total N
N=	63	38	30	8	139

Sex	Male	Female	Missing	Total N
N=	116	20	3	139

Company CO?	Yes	No	Missing	Total N
N=	33	102	4	139

d. Unit Performance.

The statistical plan for analyzing performance scores is described here briefly to aid in following the presentation of results. The purpose of the test was, "to assess the effects of varying the percentage of female soldiers assigned to representative types of Category II and III TOE units on the capability of a unit to perform its TOE mission under field conditions." Experimentally, the object was, "to provide empirical data to test the null hypothesis that specified increases in the proportion of women in selected TOE units will not impair unit performance."

An average score was obtained for each ARTEP by adding the individual overall scores for the critical tasks and dividing by the number of tasks actually scored. The major statistical analyses focused on the twice-tested companies, the experimental group and the control group. To test for a practice effect from repeated testing, using the five companies in the control group, difference scores were computed by subtracting each company's second score from their first score. These five differences scores (one from each type of company) were then used to compute a correlated observation t-test. Difference scores also used to test the effect of going from 0% women to 15% and the effect of going from 15% to 35%. In each case, a t-statistic was computed and compared to the tabled t- value for four degrees of freedom (for $p<.05$ and 4 df, $t=2.78$). In all of the above analyses, the difference score was obtained from the same company. To test the significance of the difference in performance between the companies with 0% women and those with 35%, a group comparison t-test was used since different companies were involved at the two levels of fill.

(1) Control Group Comparisons.

The first comparison, between the first and second testing of the control companies is presented in Table 16. The difference scores, as stated above, were obtained by subtracting the second score from the first score. In four out of five cases, the second score was lower than the first, a finding which will be discussed later in this part. The t-statistic revealed no significant difference in the two sets of scores (p>.05). It will be recalled that the control companies were tested first at whatever percentage of women they had and that the company personnel were to be stabilized, as much as practicable, and tested the second time with approximately the same percentage of women. As can be seen, there were some changes in the percentages of women on the two exercises for individual companies, but they were roughly comparable. Finally, no significance should be attached to the differences in average scores between types of companies. As mentioned previously, different teams of evaluators rated the different types of companies using scenarios, tasks, and scoring modules unique to each of the five unit types. There was no way to insure comparability of rating standards among the rater teams. There was, however, continuity within teams, so that in most cases the same evaluators scored both the first and second test, and the same scenario and tasks were used both times. The few exceptions to the planned continuity of evaluators as an experimental control have already been noted.

TABLE 16

AVERAGE PERFORMANCE SCORES
(Control Group)

	FALL		SPRING		
Type of Company	% Women	Mean Score	% Women	Mean Score	Difference Score
Maintenance	9.03	2.61	9.80	2.79	- .18
Medical	24.49	2.51	21.57	2.08	+ .43
Military Police	8.3	2.11	11.70	1.97	+ .14
Signal	24.07	2.13	10.29	1.85	+ .28
Transportation	0.00	2.45	0.00	2.41	+ .04
Average	13.178	2.362	10.672	2.220	+ .142

t= +1.37, p>.05

(2) Experimental Group Comparisons.

The three comparisons for the experimental group are presented in Tables 17, 18 and 19. Table 17 shows that, on the average, there was a very slight and statistically non-significant decrement in average score with an increase from 0% to 15% EW. The percentage of women in the field in all cases was close to the target of 15%. Although four out of five companies showed a slight decrement on the second test, the Maintenance Company improved their score. Table 18 presents the data for companies that went from 15% to 35% EW. There was a slight, and again non-significant, improvement in scores from the first to the second test. Finally, Table 19 presents the data for the comparison between the five companies tested first at 0% EW and the five tested second at 35%. This group comparison shows a slight and non-significant decrement in average score on the second test at the higher percentage of women. Following a method for combining independent results to obtain one overall probability, the t-statistics from the experimental group comparisons were converted to exact probabilities and then to chi-squares, each with two degrees of freedom. The chi-squares were then added and the resulting value with 6 df compared to the tabled value to determine the probability of obtaining a similar chi-square statistic by chance. The resultant combined chi-square, with 6 df, was 4.74, $p > .70$ and non-significant.

TABLE 17

AVERAGE PERFORMANCE SCORES
(0% - 15%)

	FALL		SPRING		
Type of Company	% Women	Mean Score	% Women	Mean Score	Difference Score
Maintenance	0.00	2.06	16.20	2.37	- .31
Medical	0.00	2.27	17.65	2.26	+ .01
Military Police	0.00	1.97	14.30	1.77	+ .20
Signal	0.00	1.97	12.71	1.87	+ .10
Transportation	0.00	2.68	17.00	2.59	+ .09
Average	0.00	2.19	15.572	2.172	+ .018*

*$t = +.206$, $p > .05$

TABLE 18

AVERAGE PERFORMANCE SCORES
(15% - 35%)

	FALL		SPRING		
Type of Company	% Women	Mean Score	% Women	Mean Score	Difference Score
Maintenance	16.58	1.68	35.78	2.26	- .58
Medical	18.33	2.01	37.50	2.10	- .09
Military Police	11.70	1.90	26.90	1.97	- .07
Signal	16.13	2.07	35.71	1.90	+ .17
Transportation	22.00	2.23	34.78	2.41	- .18
Average	16.943	1.978	34.134	2.128	- .150*

*t = -1.23, p > .05

TABLE 19

AVERAGE PERFORMANCE SCORES
(0% - 35%)

	FALL		SPRING		
Type of Company	% Women	Mean Score	% Women	Mean Score	Difference Score
Maintenance	0.00	2.06	35.78	2.26	- .20
Medical	0.00	2.27	37.50	2.10	+ .17
Military Police	0.00	1.97	26.90	1.97	.00
Signal	0.00	1.97	35.71	1.90	+ .07
Transportation	0.00	2.68	34.78	2.41	+ .27
Average	0.00	2.19	34.134	2.128	+ .062*

*t = +.777, p > .05

A further analysis was conducted by considering all eight companies of each type. Using the first cycle test for the experimental and control groups, and both first and second cycle tests for the calibration, once-tested group, the eight companies were divided into two groups; those with the lowest level of fill and those with the highest level of fill of women. Within each type of company, simple t-tests were computed, coverted first to exact probabilities and then to chi-square. Table 20 presents the results of this analysis. The combined X2 was 7.618 with 10 df, p>.70. Combining these X2s with those obtained earlier, a value of X2 = 12.358 with 16 df was obtained, p>.80.

TABLE 20

AVERAGE PERFORMANCE SCORES
(Low vs High Fill)

Company Type	Low Fill	High Fill	t	Pr	X^2
Maintenance	2.23	2.18	+.249	.41	1.784
Medical	2.17	2.15	+.170	.44	1.642
Military Police	1.82	1.88	-.286	.61	.988
Signal	1.97	1.965	+.055	.48	1.468
Transportation	2.41	2.38	+.219	.42	1.736
					Total X^2 = 7.618*

*p >.70

To better visualize the major findings, Figures 1 through 5 present average scores plotted against the percentage of women in the field during the ARTEP. The two points representing the two tests of the experimental companies have been connected by a line. An arrow added to the lines for the control companies indicates the temporal order of testing. The temporal order for the experimental companies reads from left to right. The unconnected points represent the results for the calibration, once-tested companies. With the exception of the Military Police companies, the calibration companies demonstrate relatively little variability of mean score, regardless of the percentage of women. All five graphs, considered as scatterplots relating the two variables, fail to reveal any consistent trends. Either these data show essentially random variations, or variables other than content of women are contributing most of the variation in performance as measured by the ARTEPs.

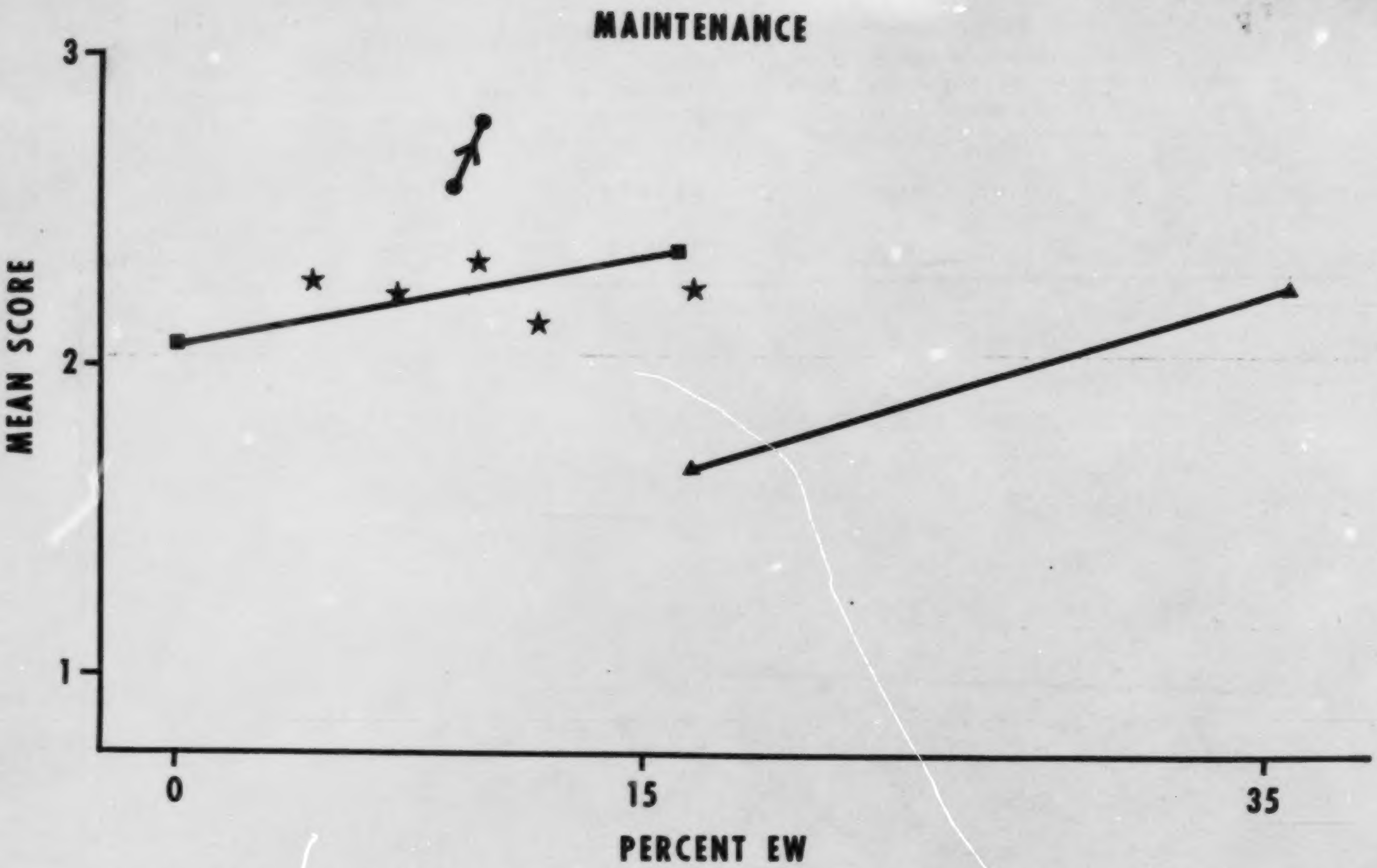


Figure 1. Average ARTEP performance scores for Maintenance companies plotted as a function of percentage of EW in the Field. Experimental Group companies (squares and triangles) and the Control Group Company (circles) have data points connected by lines. An arrow superimposed on the line for the Control Group indicates temporal order of testing. Calibration Group companies are represented by single points (stars).

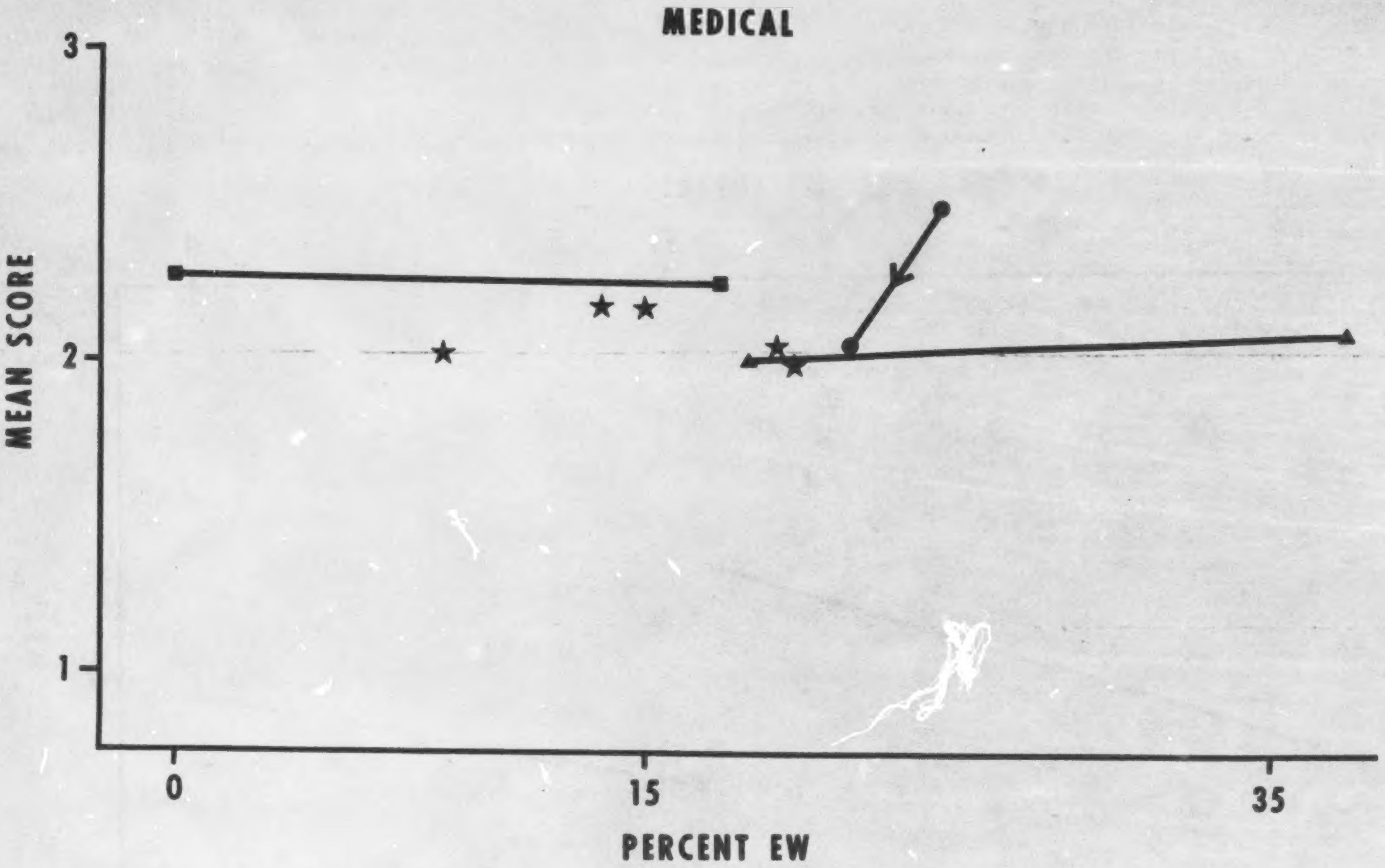


Figure 2. Average ARTEP performance scores for Medical companies plotted as a function of percentage of EW in the Field. Experimental Group companies (squares and triangles) and the Control Group Company (circles) have data points connected by lines. An arrow superimposed on the line for the Control Group indicates temporal order of testing. Calibration Group companies are represented by single points (stars).

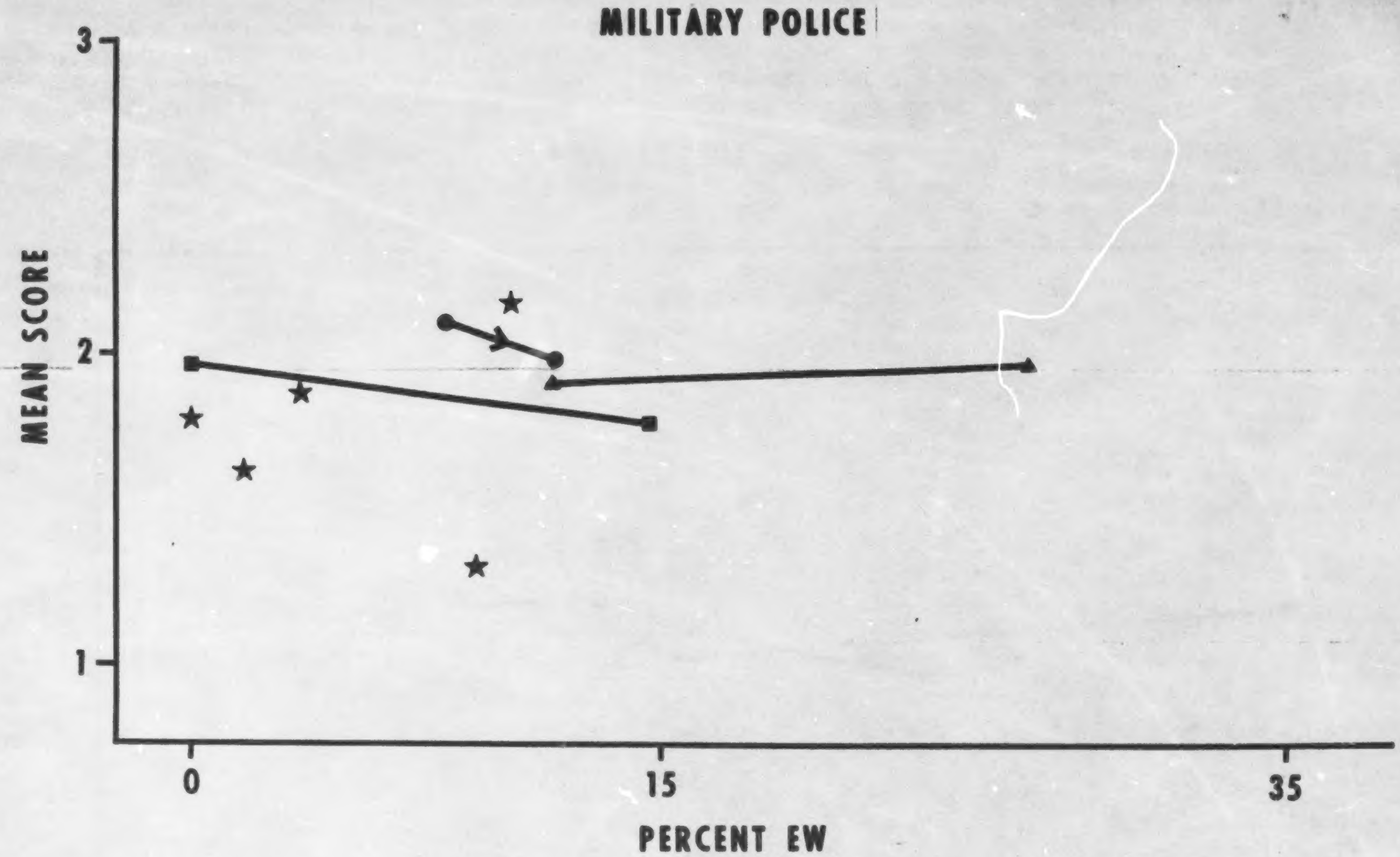


Figure 3. Average ARTEP performance scores for Military Police companies plotted as a function of percentage of EW in the Field. Experimental Group companies (squares and triangles, and the Control Group Company (circles) have data points connected by lines. An arrow superimposed on the line for the Control Group indicates temporal order of testing. Calibration Group companies are represented by single points (stars).

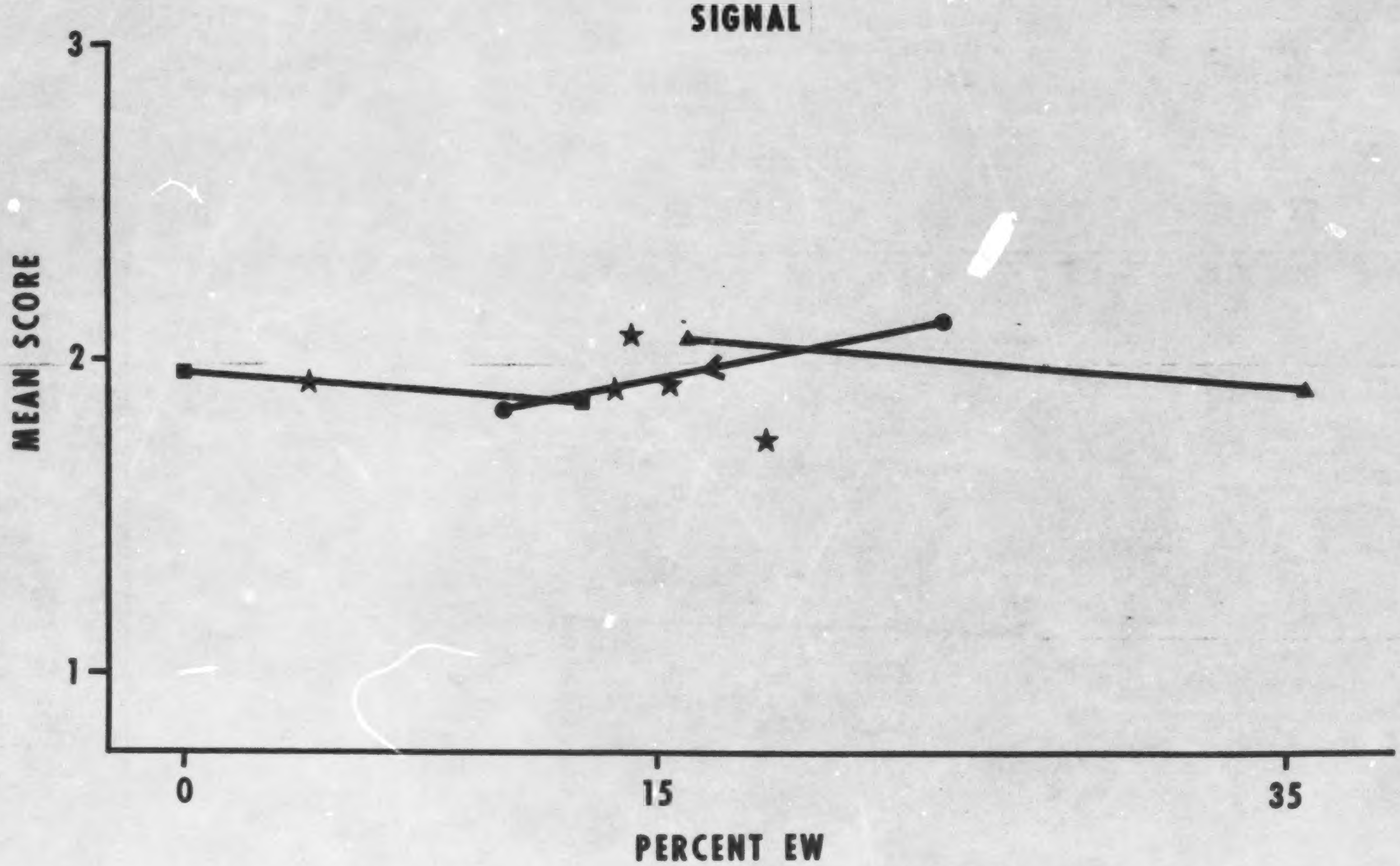


Figure 4. Average ARTEP performance scores for Signal companies plotted as a function of percentage of EW in the Field. Experimental Group companies (squares and triangles) and the Control Group Company (circles) have data points connected by lines. An arrow superimposed on the line for the Control Group indicates temporal order of testing. Calibration Group companies are represented by single points (stars).

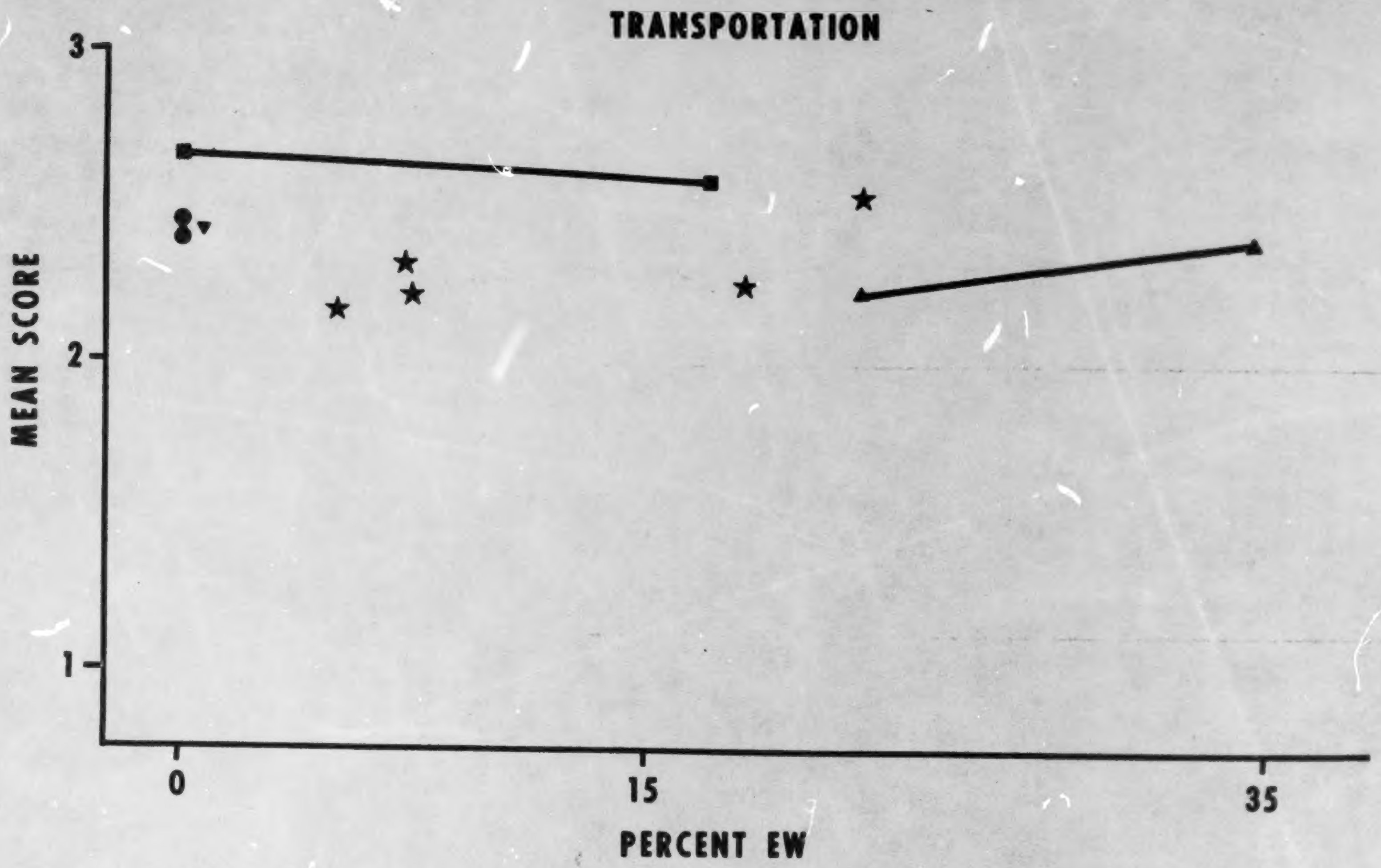


Figure 5. Average ARTEP performance scores for Transportation companies plotted as a function of percentage of EW in the Field. Experimental Group companies (squares and triangles) and the Control Group Company (circles) have data points connected by lines. An arrow superimposed on the line for the Control Group indicates temporal order of testing. Calibration Group companies are represented by single points (stars).

e. Distribution of Scores.

One possible problem in an evaluation procedure dependent on a scoring system with only three categories is that the raters might not make fine enough distinctions in assigning scores. Table 21 summarizes the overall scores awarded by each team on all ARTEPs.

TABLE 21

FREQUENCY OF SCORES

Type of Company	SCORES 1	 2	 3	Not Scored	Total N	Mean Score
Maintenance	30	200	118	92	440	2.29
Medical	102	445	215	84	846	2.15
Military Police	86	197	41	6	330	1.86
Signal	59	321	36	90	506	1.94
Transportation	17	108	117	0	242	2.41
N =	294	1271	527	272	2364	
% =	12.44	53.76	22.29	11.51	100.00	

The fact that almost 12% of the tasks were not scored probably reflects the special problems encountered with Maintenance and Signal companies. It was difficult to ensure that various repair capabilities of Maintenance companies could be demonstrated because of the lack of dead-lined equipment and the fact that some of the companies did not normally perform certain maintenance duties in garrison. The Signal companies were hampered since the ARTEPs were generally conducted as company exercises when it would have been better to evaluate the Signal companies as a part of a larger exercise to ensure adequate message traffic.

A second potential problem with the scoring system used was the possibility of the evaluator's scoring standards shifting. It will be recalled that, because there was no time for them to gain experience by running practice ARTEPs, the evaluators were instructed to maintain the same standards adopted for the first ARTEPs. To test for any systematic trends in the scores, the mean overall scores were listed in sequential order of testing. Table 22 presents these data for each company type. A simple, non-parametric runs test was conducted on the direction of change from one test to the next; i.e., to see whether the mean score went up or down. The five tests revealed only a random assortment of scores.

TABLE 22

SEQUENTIAL TEST SCORES

Order of Testing	COMPANY TYPE Maint.	Medical	Mil. Pol.	Signal	Trans.
1	2.61	2.03	1.63	2.13	2.45
2	2.27	2.01	2.11	1.97	2.50
3	1.68	2.27	1.32	2.08	2.68
4	2.33	2.18	1.97	1.74	2.55
5	2.06	2.51	1.90	2.07	2.23
6	2.29	2.19	1.87	1.90	2.32
7	2.79	2.26	1.80	1.92	2.23
8	2.23	2.08	2.17	1.85	2.18
9	2.26	2.06	1.97	1.93	2.41
10	2.15	2.00	1.77	1.90	2.59
11	2.37	2.10	1.97	1.87	2.41

f. Secondary Criterion Measures.

The collateral research questionnaires afforded an opportunity to ask those involved in the exercises to assess their own performance on the ARTEP. Consequently, both the officer and the enlisted questionnaires asked the participants to rate how well their company, how well the women, and how well the men did on the ARTEP. Additonally, the enlisted soldiers were asked to rate the performance of their squad or section and their own performance on the ARTEP.

The results for the first question, "How did your company perform on the ARTEP?", are presented in Table 23. The response categories have been collapsed for ease of presentation and the enlisted data broken down by junior and senior enlisted for both sexes. The last response category actually read, "Don't know, not sure."

The self-ratings of company performance made by the enlisted soldiers in the twice-tested companies were analyzed separately. A change score was computed by comparing the average self-rating score on the first test with that from the second test. These self-ratings were then compared to the change in evaluator scores from the first to the second test. Using the mean overall score from the evaluators vs the average self-rating score, there was agreement on the direction of change in scores from the first to the second test in 12 out of 15 cases. A X^2 test, corrected for continuity, shows this to be significant at the $p < .05$ level. Table 24 summarizes the results, while an expanded version with evaluator scores and self-rating values can be found in Table 24a.

TABLE 23

HOW DID COMPANY PERFORM?
(in %)

Responses	Officers (N=139)	MALES E1-E4 (N=3556)	MALES E5-E9 (N=1552)	FEMALES E1-E4 (N=762)	FEMALES E5-E9 (N=63)
Outstanding/Very Well	70.5	64.8	70.2	63.7	68.3
Fairly Well	27.3	27.5	24.2	28.4	23.8
Rather/Very Poorly	2.1	4.4	3.6	3.3	4.8
Don't Know	0.0	3.8	2.0	4.6	3.2

TABLE 24

CHANGE IN EVALUATOR SCORES AND SELF-RATINGS FROM FIRST TO SECOND ARTEP

Company Type or Group		Evaluator Scores*	Self-Ratings*	Agreement?
Maint.	0-15%	+	+	Yes
	15-35%	+	+	Yes
	Control	+	+	Yes
Medical	0-15%	-	-	Yes
	15-35%	+	+	Yes
	Control	-	-	Yes
MP	0-15%	-	-	Yes
	15-35%	+	+	Yes
	Control	-	+	No
Signal	0-15%	-	-	Yes
	15-35%	-	-	Yes
	Control	-	+	No
Trans.	0-15%	-	-	Yes
	15-35%	+	+	Yes
	Control	-	-	Yes

* + indicates an increase; - indicates a decrease in quality of performance

TABLE 24a

Comparison of Evaluator Awarded Scores with Self-Ratings by Enlisted Personnel

Company Type & Group	EVALUATOR SCORES Test 1	 Test 2	 Change	SELF-RATINGS Test 1	 Test 2	 Change*
Maint. 0-15%	2.06	2.37	+0.31	2.43	2.13	+0.30
" 15-35%	1.68	2.26	+0.58	2.52	2.21	+0.31
" Control	2.61	2.79	+0.18	1.79	1.57	-0.22
Med. 0-15%	2.27	2.26	-0.01	1.46	1.88	-0.42
" 15-35%	2.01	2.10	+0.09	2.18	1.75	+0.43
" Control	2.51	2.08	-0.43	1.63	1.89	-0.26
MP 0-15%	1.97	1.77	-0.20	2.08	2.12	-0.05
" 15-35%	1.90	1.97	+0.07	2.18	2.09	+0.09
" Control	2.11	1.97	-0.14	2.41	2.20	+0.21
Sig. 0-15%	1.97	1.87	-0.10	2.23	2.43	-0.20
" 15-35%	2.07	1.90	-0.17	2.02	2.26	-0.24
" Control	2.13	1.85	-0.28	2.66	2.09	+0.55
Trans. 0-15%	2.68	2.59	-0.09	2.131	2.133	-0.002
" 15-35%	2.23	2.41	+0.18	2.34	1.99	+0.35
" Control	2.45	2.41	-0.04	2.02	2.42	-0.40

* A smaller value indicates a better self-rating

Tables 25 and 26 present the data from the questions asking separately how well women and men had performed on the ARTEP. The reduced Ns in Table 25 reflect the fact that only data from companies with women were used. Since the question did not direct attention only to enlisted performance, there may have been some confusion for those companies with female officers.

TABLE 25

HOW DID WOMEN PERFORM?
(in %)

Responses	Officers (N=131)	MALES E1-E4 (N=2987)	MALES E5-E9 (N=1353)	FEMALES E1-E4 (N=740)	FEMALES E5-E9 (N=61)
Outstanding/ Very well	68.00	44.8	56.2	71.9	78.7
Fairly well	19.8	31.8	28.5	21.1	18.0
Rather/Very Poorly	12.3	13.7	10.4	2.8	1.6
Don't know	0.0	9.8	5.0	4.2	1.6

TABLE 26

HOW DID MEN PERFORM?
(in %)

Responses	Officers (N=138)	MALES E1-E4 (N=3589)	MALES E5-E9 (N=1563)	FEMALES E1-E4 (N=762)	FEMALES E5-E9 (N=63)
Outstanding/ Very well	79.0	72.2	75.8	70.0	76.2
Fairly well	18.1	23.4	20.9	23.1	20.6
Rather/Very Poorly	2.9	2.1	1.9	3.4	1.6
Don't know	0.0	2.1	1.3	3.5	1.6

Finally, Tables 27 and 28 present the data from the questions asking how well the respondents thought their own squad or section had performed and how well they thought they had performed. Three observations can be made at this time about these data. First, the opinion of more than 10% of the officers and EM that women performed "rather" or "very poorly" was not shared by female enlisted ($X^2 = 56.68$, $p < .001$). A second observation concerns the opinions of all enlisted groups about the performance of their own squad or section. The frequently substantiated observation about the importance of the soldier's immediate comrades is borne out by the generally high ratings given by all enlisted groups to his, or her, squad or section. Finally, it would seem from Table 28 that senior enlisted males have the highest opinion of their own performance and the lower ranking females had the lowest.

TABLE 27

HOW DID YOUR GROUP PERFORM?
(SQUAD OR SECTION)
(in %)

	MALES		FEMALES	
Responses	E1-E4 (N=3595)	E5-E9 (N-1563)	E1-E4 (N-761)	E5-E9 (N-63)
Outstanding/Very well	78.9	84.2	77.4	81.0
Fairly well	17.0	13.2	16.6	17.5
Rather/Very poorly	2.6	2.2	4.4	1.6
Don't know	1.6	.4	1.6	0.0

TABLE 28

HOW DID YOU PERFORM?
(in %)

	MALES		FEMALES	
Responses	E1-E4 (N=3595)	E5-E9 (N=1563)	E1-E4 (N=758)	E5-E9 (N=62)
Outstanding/Very Well	69.8	79.9	63.3	72.6
Fairly well	25.7	17.3	31.3	21.0
Rather/Very poorly	2.4	1.9	3.9	6.4
Don't know	2.1	1.0	1.5	0.0

g. Factors Affecting Unit Performance.

At the conclusion of the fall testing cycle, members of the Test Directorate expressed the view, both collectively and individually, that even though they had not observed companies with 35% women, they felt that variables other than the percentage of women were more important in determining unit performance. As a result of a number of discussions about their first-hand observations, a question was constructed for the officer's questionnaire, then being developed. Essentially, it asked the officers to consider five factors which may affect a company's ability to carry out its mission. They were then asked to apportion 100 percentage points to these five factors (plus an open-ended sixth factor if they wished to add to the list) according to the degree they thought the factors contribute to a company's real ability to accomplish its mission. Although admittedly hypothetical, the consistency of results merits its inclusion in this report. Table 29 shows how the 134 officers answering the question apportioned the 100 points among the factors. Cell entries are the percentage of respondents awarding a percentage in that range to the factor listed on the left. Where the apportionment totaled less than 100 points, a statistical correction was made.

TABLE 29

FACTORS CONTRIBUTING TO A COMPANY'S CAPABILITIES
(in %)

	Apportioned Percentage Points										
Factor	0	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
Leadership	.7	3.7	21.6	34.3	17.9	17.2	2.2	.7	0	.7	.7
Training	.7	10.4	17.9	34.3	20.1	12.7	3.7	0	0	0	0
Morale	5.2	17.2	47.8	22.4	4.5	2.2	.7	0	0	0	0
Personnel Turbulence	14.2	58.2	23.1	3.0	1.5	0	0	0	0	0	0
% Women	24.6	63.4	10.4	.7	0	0	.7	0	0	0	0
Other	82.1	11.9	4.5	0	.7	0	.7	0	0	0	0

Table 30 summarizes these data, showing for each factor the median value, the mean value, and the interquartile range. The latter summary statistic indicates those values comprising the middle 50% of the values and is a measure of the dispersion of values. As can be seen, the distributions are relatively tight, indicating a fairly strong consensus regarding the relative importance of these factors in affecting a company's ability to perform its mission.

TABLE 30

SUMMARY OF FACTORS CONTRIBUTING TO A COMPANY'S CAPABILITIES

Factor	Median % Value	Mean % Value	Interquartile Range
Leadership	30	32.119	19-37
Training	30	29.687	19-37
Morale	20	19.612	13-23
Personnel Turbulence	10	9.754	5-13
% Women	5	6.687	0-10
Other	0	2.164	0

2. DISCUSSION

a. Introduction.

Some of the problems in conducting the present test have been identified and discussed in Part II of this report. Further discussion of some of them is merited in light of the independent analysis of the test made by the Operational Test and Evaluation Agency (OTEA) at the tasking of the Director of the Army Staff.

b. ARTEP Validity.

The Army Training and Evaluation Programs are the product of service schools and contain the tasks considered critical for the accomplishment of a unit's TOE mission. By TRADOC doctrine, the ARTEP provides guidance for a company commander to construct 3-day training exercises as a means for diagnosing the training needs of the unit. The document does not dictate a particular scenario to be used in conducting the exercise but provides guidance for choosing tasks to be included in a comprehensive assessment. Although some of the ARTEPs were in coordinating draft form at the inception of the project, the stated opinion of those involved in producing them was that there would be few changes (mostly minor) when published as Test Editions. They were, in other words, very close to being operational and ready to be sent to the field. The ARTEPs were not developed experimentally, nor were they developed specifically for use in the present project. As previously discussed, the ARTEP is the

DA approved instrument for measuring unit performance for the purpose of identifying specific training needs. They were developed by the branch schools, making use of existing ATTs and their own resident expertise. The ARTEPs, though not designed as tests per se, are the official means of evaluating a unit's capabilities. It should be noted that ARI received special permission from TRADOC for the one-time use of ARTEPs as performance tests in this project.

The questions included in the collateral research questionnaires about the ARTEP take on added significance because these exercises were, in many cases, the first time a unit was evaluated using the newly developed ARTEPs. The positive response of a large majority of those participating in the exercises lends credence to the view that the ARTEPs constituted realistic tests of the companys' ability to perform its military mission.

c. Selection of Participant Companies.

It was not possible to randomly select companies from CONUS installations for assignment to the project. In some cases, the need for eight companies with the same TOE almost exhausted the number available. However, the personal background information, e.g., age, education, presented earlier would indicate that the soldiers participating in the project are representative of the Army as a whole. Women are concentrated in the lower enlisted grades, are slightly older than their male peers, are better educated, are less likely to be married and probably, are a bit taller and heavier than their civilian counterparts.

d. Control Group Companies.

The control group was included in the research design to assess the effects of a company being tested twice. The concern here was that a company would "learn from its mistakes" and improve on the second test. Table 16 showed that four out of five companies actually had lower scores on the second test. A possible explanation for this finding was the fact that the first ARTEP counted as "official" for these companies, while the second ARTEP was conducted solely for the purposes of the project. The unofficial nature of the second ARTEP was also true for the experimental companies. It should be noted that, in the middle of the project, DA eliminated the requirement of an annual ARTEP for these companies.

e. Test Scores.

Statistical comparisons for the experimental group failed to reveal any significant differences related to percentage of women. Level of female fill was not systematically related to unit performance if all 55 ARTEPs are considered. Many of the questions that were raised after the start of the project, although of great interest to the Army, are not germane to the issues addressed by the present test. For example, while it may be fruitful to ask, in retrospect, whether three days is sufficient to test the capabilities of women, since a three day exercise was specified in the charter for the project this question suggests an alternative which is entirely outside of the scope of the test.

Table 21 presented the distribution of task scores awarded by the evaluator teams. As is evident from Table 21, two of the teams (MP and Signal) tended to award lower scores than the other three. Without an independent evaluation of the companies, it is impossible to tell whether there were true differences between types of companies or whether the differences simply reflect different scoring standards of the teams. Examination of sequential mean overall scores (Table 22) revealed no pattern which might suggest changes in scoring standards. The percentage of "3's" is not especially higher than it should be according to the instructions given to the evaluators. The number of unscored tasks, however, was disappointingly high. If more time had been available to develop the scenarios and to pilot test procedures, some tasks would probably have been eliminated and others substituted because of the probability that particular events could not be scheduled for all companies. Exigencies at the installation level resulted in some companies being structured differently than specified in the TOE. Additional preparation time would likely have surfaced these problems and would have permitted changing the scenarios accordingly.

f. Collateral Measures.

(1) Collection of the opinions of the respondents/participants about their own performance was deemed an important data source. For the most part, the evaluators gave "passing" grades to all but a few companies. This assessment was shared by a majority of the individuals involved in the test. Although the rank and sex breakdowns showed some disagreements, they were relatively minor. The assessment of females' performance showed the greatest lack of consensus. Females did not share the opinion of some male enlisted soldiers and the officers. Over 10% of the officers and EM felt that women had performed "rather poorly" or "very poorly." Interestingly, the more senior enlisted and the officers had a higher opinion than did the lower enlisted ranks. Also, the latter group were more reluctant to express a definite opinion with almost 10% answering "don't know, not sure." It may be significant that the lowest rating of women's performance was made by their male peers and this opinion was not shared by more senior male enlisted, or by the officers. Finally, it should be noted that women in the lower enlisted ranks gave more high ratings of the performance of their male counterparts than the males gave to them.

(2) In the course of conducting the first two dozen field exercises, the members of the evaluator teams perceived that unit performance had little to do with the proportion of women in the companies. The women observed and rated during the test were primarily AIT graduates, competent in their jobs, and motivated to do well. Recruitment standards for women were such to insure that only brighter, better educated, and slightly older women were brought into the Army during the period from 1972 until 1976 when the project was initiated. It is not surprising, therefore, that companies with even a relatively large proportion of women performed well. Most of the company officers felt that the percentage of women, per se, contributed only a minor part to the company's performance in the field. Training, morale and leadership were perceived as the major factors contributing to the company's ability to perform its mission. The inference here is that percentage of women is relatively unimportant if they are well-trained, well-led, and well-motivated to perform.

g. Control of Variables.

Part II of this report discusses the need to control variables which might affect unit performance. The attempts made to control variables were not always completely successful; however, major considerations in conducting the test included that installation policies would not be contravened by DA Washington, that career advancement would not be hampered by participation in the test, and that there would be no compensation for adverse weather.

The twice-tested units belonging to the experimental and control groups were to have the commanders stabilized. This was accomplished in all but two cases. A Signal company in the control group had a change of command between the two tests. The MP company commander of the unit which went from 0% to 15% was promoted to 04 and transferred. All other repeated testing was conducted with the same company commander.

There was more personnel turbulence than planned in three control companies (MP, Signal and Transportation) and in two of the experimental companies that went from 15% to 35%. Some of the once-tested companies experienced more personnel turbulence than planned, and approximately one third had 10% more turnover during the 60 days prior to the ARTEP than specified.

Weather was generally favorable for most of the experimental and control group tests. Maintenance and Transportation companies experienced no adverse weather on any of their exercises. Two Medical and two MP companies experienced adverse weather (rain and high winds or extreme cold or snow), as did one Signal company (rain, snow and sleet). One test was cancelled and rescheduled because of sub-zero temperatures.

Attainment of the proper female fill was particularly difficult for those companies with the highest proportion of women. In at least four cases, experimental companies did not have the full 60 days to prepare for the ARTEP with all personnel available for duty. Two tests were postponed to allow a minimum of 30 days preparation for the ARTEP. The 60-day period specified in the OTP was chosen to allow sufficient time for (1) people to get acquainted and (2) training for the exercise. It was recognized, at the time the OTP was prepared, that it would be necessary to cross-fill using installation personnel resources to attain the desired proportion of women with the proper distribution of MOSs. This necessarily meant temporary assignments in many cases. It was believed that 60 days was enough time for MOS-qualified women and men to become accustomed to the unit, its officers and NCOs. Relaxation of the 60-day requirement was made for two reasons. First, firm deadlines for reporting the results necessitated finishing all tests by the end of June 1977. Second, since all enlisted women assigned to these units had to be MOS-qualified, it was felt that 60 days was a generous estimate of the time necessary to become (if only temporarily) assimilated into the unit, especially since many were already assigned to the post.

The problems of only 30 days preparation time occurred almost exclusively with the companies that went from 15% to 35%, and it is instructive, therefore, to review Table 18 which presents the mean scores for these companies. The average overall scores for four out of five of the companies were higher on the second test. Although a statistical test on difference scores fails to show a significant change, the failure to adhere strictly to the controls specified in the OTP did not appear to greatly affect the results in the expected direction.

There were some posts where post policy influenced the use of enlisted women, although it is doubtful that overall company scores were affected. Several examples serve to illustrate this influence. Three of the posts required that enlisted women sleep in a common tent, which probably altered the normal (i.e., all male) deployment of soldiers in the bivouac area. Additionally, one of these posts required that women move only in pairs after dark. Although the post policy in this case was the result of a rape/murder several years before, most enlisted women were unaware of the basis for the policy and expressed resentment at the differential and deferential treatment.

3. CONCLUSIONS

This research project was designed to examine the hypothesis that specified increases in the proportion of enlisted women in selected TOE units will not impair unit performance. The evidence presented here indicates that the hypothesis, given the parameters studied, cannot be rejected. In plain language, the data indicate that proportion of women, up to the percentages studied, had no effect on measures of unit performance in the field. In the course of conducting this project, many issues concerning the utilization of women in the Army have surfaced. Some of these issues, such as the physical strength and stamina of women, may be studied objectively in separate studies. Others, such as the advisability of placing women in situations likely to involve them in actual combat, can only be partially answered by research. The likelihood of unit contingency missions involving support units in combat can be assessed by simulation or war games and the performance of women soldiers in simulated tactical situations can be evaluated, but the impact that a large casualty rate among women would have on the American public has to remain a subjective judgment. A valid answer to this question cannot be obtained in an opinion survey. Integration of increasing numbers of women into non-traditional jobs in the Army is only beginning. There is anecdotal evidence from the project and elsewhere that resistance to women soldiers tends to abate when males have first-hand experience working with them. It takes time, however, and total acceptance is not just around the corner.

PART IV

ARI INTRODUCTORY REMARKS RE CONTRACTOR ANALYSIS OF TEST DIRECTORATE TEAM OBSERVATIONS AND EVALUATIONS

WOMEN CONTENT IN UNITS

1. The Test Directorate evaluator teams completed comprehensive after-action reports for each exercise. These consisted of a package of materials including the scoring sheets, basic supporting documents such as maps, Unit Manning Reports, copies of messages, and a memorandum summarizing pertinent observations about the exercise as a whole. The latter was written under general guidelines that it report certain specified observations such as terrain, road trafficability, etc., and that it should also contain the evaluators unrestrained reactions to the conduct of the exercises, any problems encountered, the kind of support given them by the installation, and any other comments which might aid in interpreting the data collected on the exercise. The teams were encouraged to comment freely on any aspect of the exercise they deemed important or significant.

2. The original plan was to have the Test Directorate teams provide an overall summary of their observations as embodied in the after-action reports and their own experiences gained from almost a year-long involvement in the project. This proved impracticable at the end of the project because many of the officers had to either return to their assignments or report to new assignments. All of the teams did have time, however, to pool their collective experience and to comment on, in response to a request from the Test Director, some hypotheses drawn up about the role and utilization of women.

3. It was decided to subject these sources of data to an independent analysis by outside scientifically sophisticated analysts. Consequently, a contract was let for a firm experienced in behavioral science research to study and analyze the after-action reports, the hypothesis file, and other source documents recording the experience of those conducting or observing the ARTEPs, and to report their findings. It should be noted however, that the conclusions stated in that report represent the opinions of the contractor based on his study of the data sources mentioned above.

TECHNICAL REPORT NO. 111

QUALITATIVE ANALYSIS OF SUBJECTIVE EVALUATION OF WOMAN CONTENT IN UNITS (MAX-WAC) FTD

THOMAS C. WYATT

JOHN F.C. KENNEY, JR.

A.M. ROBERT DEAN

PREPARED UNDER CONTRACT DAHC19-77-M-0045
FOR THE U.S. ARMY RESEARCH INSTITUTE FOR
THE BEHAVIORAL & SOCIAL SCIENCES
5001 EISENHOWER AVENUE
ALEXANDRIA, VIRGINIA 22333

22 SEPTEMBER 1977

1. INTRODUCTION

1.1 General. This section presents a qualitative analysis of subjective evaluations by US Army Research Institute personnel of Women Content in Units (MAX-WAC) field tests conducted during the period October 1976 through June 1977. This analysis was performed by contract personnel who are trained in scientific research methods. Therefore, the analysis benefits from (1) the absence of subjective association with the test agency, and (2) applied knowledge of evaluation research methodology.

1.2 Analysis. Analysis was performed on subjectively arrived at findings and conclusions on the effect of the presence of women soldiers in five types of Army units. Data were drawn from reports identified in three categories.

1.2.1 Test Directorate Team Reports

1.2.2 Army Research Institute (ARI) Staff Visit Trip Reports

1.2.3 Hypotheses constructed from ARTEP observations.

1.3 Findings. Findings are presented for each type unit by each data category.

1.4 Conclusions and Recommendations. Conclusions based on interpretation of the findings complete this section of the report. Recommendations are not considered appropriate for this section of the report.

2. METHODOLOGY

2.1 Description of Data Used. Subjective evaluation data was drawn from the following reports.

2.1.1 Test Directorate Team Reports

2.1.2 ARI Staff Visit Trip Reports.

2.1.3 Hypotheses Constructed from ARTEP Observations.

2.2 Test Directorate Team Report. Team members of the MAX-WAC Test Directorate, described earlier in this report, visited each unit selected to participate in the field tests and observed each unit test. These observations were, generally, subjective assessments of training/test areas, personnel status, unit organization and structure, overall impression of unit performance, under varying conditions, and observations of activities or situations peculiar to the tested unit or of special interest to the evaluator.

2.2.1 There was consistency of report format within type of unit (MP, Medical, etc.), but not across the various types; e.g., the report format used for MP units was different from that used for medical units.

2.2.2 A total of fifty five (55) Test Directorate Reports was analyzed.

2.3 ARI Staff Visit Trip Report. Selected members of the ARI assigned staff visited ten units scheduled for the field tests. Units visited included three Maintenance, three MP, two Signal, and one each Medical and Transportation type companies. All visits were made during the early part of the field test phase. Information obtained during these visits was used as a substitute for a pilot test which could not be scheduled. Information contained in these reports is described as subjective assessments and observations of the administrative problems which could be encountered in the future.

2.4 Hypotheses Constructed from ARTEP Observations. In late May, 1977, the Test Directorate formulated a total of fifty eight (58) statements of experience relating to the utilization of Army personnel, male and female, based on ARTEP observations. These hypotheses, represented the tentative assessment of evidence collected during the on-going field tests. Each of the five Test Directorate Teams (MP, Medical, Maintenance, Transportation, and Signal) was tasked to address each "hypothesis" with a synopsis of relevant observations, presenting discussions to support or to refute each hypothesis based on their test experiences. Conclusions and recommendations associated with their discussions were also to be made by Team members.

2.5 Data Sources. All data sources used in this analysis have been described above. Supporting documentation has been excluded from this section of the report due to its voluminous and, in some cases, draft style nature. All refrences are available for inspection at ARI document storage facilities.

2.6 Data Analysis. The principal approach used in this analysis was a modified form of content analysis, utilizing independent analysts each acting alone, thereby protecting against the possibility of one influencing the other.

2.6.1 First, reports in each category (Test Directorate, Trip, Hypotheses) were read, noting observations whose frequency transferred them into most frequently appearing statements. This was done within report categories, and findings recorded.

2.6.2 Next, these same data sources were examined by type of units, both within and between data categories. For example, Test Directorate Team Reports were separated into MP, Maintenance, Medical, etc. All MP units were examined, using the Test Directorate Report, then the Staff

Visit Trip Report, and then the Hypotheses. Next, a comparison was made between and within unit types according to their role in the test design — Experimental, Control, or Calibration.

2.7 Nature of the Data. Interpretation of the findings presented in this report is guided by the nature of the data which produced these findings.

2.7.1 The data analyzed in this section are comprised of a collection of subjective judgements based on observations of real events and activities. The individuals making such judgements bring into play their own personal experiences, which tend to shape their choice of what to observe, the assignment of meaning to what is observed, and the evaluation of that information. Finally, these perceptions and judgements are individually tuned and, therefore, when many observers are involved, some variability between judgements can be expected.

2.7.2 These data were generated by observers not trained in the rigorous scientific method of participant observation or nonobtrusive measurement. However, knowledge based upon experience, and applied to the interpretation of data is valuable. This is especially true when the data are related to special skills or activities, such as military operations.

2.7.3 When the individual observations are being consolidated, the assembler of these subjective judgements must make another subjective evaluation regarding the validity of the weighting and interpretation of these data into the conclusions and recommendations found in the various reports.

2.7.4 There is always some risk of error in interpreting the subjective evaluations of others. The potential for error is increased when the interpreter makes the assumption that his view of the world is the proper frame of reference from which the subjective data are to be viewed. Also, the magnitude of the error can be increased in two ways — when the interpreter is untrained; or, when well trained, assumes his training qualifies him to adopt an unchallengeable position of "best" interpreter.

3. FINDINGS

3.1 General. Findings are presented for each data category by type of unit (MP, Medical, Signal, etc.). Findings are not summarized as they are themselves summaries of information contained in the data. Therefore, some repetition will be found. This reflects the frequency and source of the information. These reported findings serve as a basis for the conclusions which follow.

3.2 Test Directorate Team Reports

3.2.1 General. The training environment was considered and reported for each unit tested. This consisted of weather, terrain, and trafficability information. The omission of observations of this nature from this analysis is based on the assumption that Army units are organized and equipped to operate in all weather and terrain, except in extremely adverse conditions. All tests were conducted in moderate, though at times disagreable, weather and terrain conditions. In general, environmental conditions did not play a part in arriving at a determination with respect to the impact of assignment of female soldiers on unit performance.

3.2.2 Signal Units.

3.2.2.1 Personnel. The average participating unit was manned at 101% of its authorized TO&E. Present for duty in the field average strength was 79% of the assigned personnel. Proportion of women in units was within test design limits.

3.2.2.2 General Evaluation. Test scenario was followed by only four of the eleven units tested* When the scenario or schedule was not followed, it was because of a lack of TO&E equipment, shortage of MOS qualified personnel, or total personnel in the field. For example, one company had only 65 percent of its equipment in the field. Although two units reported that successful pre-ARTEPS training was conducted, five units had little or no field experience operating in their TO&E mission assignments.

---

* ARI Comment: Minor variations from the scenario were permitted by the OTP to adapt to local conditions. Failure to follow the scenario to the letter resulted from conditions at the local installation. The occurred more often with Signal units than the other types of companies.

3.2.2.3 Tactical. Commanders and higher headquarters have deemphasized local security training for Signal units, assuming it will be provided by co-located non-Signal troops. Only one of the eleven units performed satisfactorily during tactical phases; that one unit was a repeated measures unit which had not performed well on the tactical phase during the first ARTEP. None of the units displayed adequate field experience, adequate training, or motivation. For example, neither work nor play was interrupted by aggressor attacks. Noise of generators prevented members hearing unit alarms for attack.

3.2.2.4 Integration of women into units. In most cases, women were not newly assigned to the units. Women displayed high morale, were accepted as equals by work peers and first line supervisors, and performed satisfactorily in team situations. Generally, females experienced problems performing tasks requiring great individual physical strength. On tasks requiring above average female strength, women would be augmented by men, or perform the task over a longer period of time. In some instances, males would not wait for the women to perform the task and would take over. Women expressed the opinion that they could perform 95 per cent of all physical tasks in the unit; an exception being, for example, starting a cold 10 KW generator by hand. Most female soldiers want field training, and need it. Most expressed objection to the requirement that they be separated from other team members for sleeping, and that they be escorted after dark. The higher the percentage of women in a unit the less pampering was observed the more the women were treated as equals. Traditional sex role definitions and expectations appear to be greatest obstacle to integration of women in units. When the chain of command expresses its attitude, negative or positive, regarding women in units, this attitude is reflected by the unit members.

3.2.2.5 Conclusions. ARTEP was well received and considered a good opportunity to train in TO&E mission assignments. However, repeated use of the same training areas detracts from the realism of the ARTEP. All members of the unit, male and female, need training in basic military skills, tactics, field exercises. No degradation of unit performance was noted by the integration of female soldiers into the unit.

3.2.3 Transportation Units

3.2.3.1 Personnel. The average participating unit was manned at 106% of the authorized TO&E. Present for duty in the field average strength was 86% of the assigned personnel. Proportion of women in units was within test design limits.

3.2.3.2 Training Status. Six of the eleven units tested reported that post support requirements interfered with training for TO&E missions. Other units did not report on that point. These same six units participated in pre-ARTEP training; three units did not, and two had pre-ARTEP training interrupted by post support requirements. All units considered ARTEPS to be a good training opportunity. Eight units reported very high personnel turnover, ranging from 39% to 106% within a one year period.

3.2.3.3 Subjective Comments. While the female soldier is usually technically qualified in her MOS, she is often deficient in basic military skills and field experience. However, these deficiencies are not revealed when evaluations in the field are only for short periods. Greatest problem areas are lack of individual physical strength and requirements for separate facilities, including hygiene and field sanitation measures. Another problem is that females are not allowed to operate alone, as are males. Acceptance of females into units cannot be legislated; they will be accepted on their individual merit. The attitude of the chain of command and higher authority can facilitate or obstruct the acceptance of women in units. Previously all-male units will not readily accept females without some prior conditioning and training. Female soldiers have higher entry qualifications than males, but the Army is not now prepared to utilize this to its advantage. Regarding task performance, with no prior civilian experience and equal military training and experience, male and female performance is about equal. An often stated objection to females in units is based on the assumption that women will be assigned or will seek traditional female roles and leave male members overburdened, or unit tasks unfulfilled. The traditional view of sex role differences encourages males to be protective of females. This is sometines exploited by the female, but not always knowingly. Further, Department of the Army policy guidance is not available to the local commander as to the proper management of female soldiers in units. Also, the present policy of assignment restrictions based on geographical limits to the rear of the brigade boundry, threatens to deny the command flexibility in the utilization of women assigned to the unit which sometimes operates in that area. If the women would have to be replaced at the last minute in order to meet this requirement, the unit would be rendered ineffective and, in turn, the combat effectiveness of the supported unit would be lowered.

3.2.4 Medical Units

3.2.4.1 Personnel. The average participating unit was manned at 93% of its authorized TO&E. Present for duty in the field average strength was 87% of the assigned personnel. Proportion of women in units was within the test design limits.

3.2.4.2 General Evaluation. ARTEP plans were not consistently followed due to resistance from local commanders who seized opportunity to conduct on the job training not included in the scenario. Female members were in some cases disproportionately assigned to sections within the test units. For example, ambulance sections were sometimes observed to be 40% female; when women experienced difficulty performing strength related tasks (loading and unloading litter patients), males were drawn from other sections to assist. Eventually, leaders began shifting females away from strength related tasks, or overloaded these tasks with males, as time in field increased. Morale was high in all units even though they were not experienced in field operations. A noticable deficiency was that personnel, male and female, lacked TO&E mission skills because training time was consumed by post support missions. Also, damaged, inoperable, or missing equipment adversely affected unit performance. Female performance was regarded as satisfactory or excellent, except for basic military skills and performance of field duties*. However, these deficiencies were also observed among male members of the units.. Many members claimed they were not used to being tested on their field medical skills, i.e., bandaging, taking and processing X-rays, changing dressings, mass causalty treatment, etc. Many were unaccustomed to the role playing associated with test and therefore uncertain as to performance expectations. Field operations continually improved with added time and experience.

---

* ARI Comment: The phrasing of this sentence may lead to some misunderstanding. ARI translates this to mean that non-MOS related duties were less well performed.

3.2.4.3 Tactical Operations. The general impression was that the medical units which were tested are unskilled in field tactical operations due to lack of training, experience and perceived need to be trained in non-MOS related skills. Road marches ranged from poor to good, and responses to aggressor action was usually poor. Organization for defense was most often unsatisfactory.

3.2.5 Maintenance Units

3.2.5.1 Personnel. The average participating unit was manned at 107% of its authorized TO&E. Present for duty in the field average strength was 84% of assigned personnel. Proportion of women in units was within the test design limits.

3.2.5.2 Pre-ARTEP Coordination. Five of the eleven units reported satisfactory cooperation and planning by higher headquarters and relief from some post support missions in preparation for ARTEP. The remaining six companies experienced poor planning, lack of cooperation, and little relief from post support missions. Filler personnel, male and female, were assigned just prior to the field test. Equipment shortages and deviations from TO&E organization were not corrected prior to movement to the field. Little or no tactical or MOS related training was conducted prior to the conduct of the test.

3.2.5.3 Tactical. Road march operations were usually good, including reaction to aggressor ambush. Movement into the bivouac area was poor. The preparation, execution, and supervision of defense operations was poor to fair due to lack of training and experience. Females participated in the tactical operations of their units and performed as well as male counterparts. Weapons training deficiencies were noticeable.

3.2.5.4 Organizational Structure. None of the units in the field were structured, equipped, or manned according to their TO&E. Their organization reflected instead their individual tailoring for post support operations. This deficiency was underscored by the lack of TO&E MOS positions and skills. One of eleven units had trained for TO&E missions. In most of the observed units, females were well integrated into units as work unit team members. Only in isolated cases were women assigned to jobs outside their MOS or given no tasks to perform. During tactical operations, women performed MOS tasks while males manned the perimeter.

3.2.5.5 Automotive Maintenance. In most cases, wheeled vehicle maintenance support was satisfactory, whereas tracked vehicle support was not, due to shortage of personnel with MOS skills or equipment. This shortage was due to the influence of a post support mission which did not include tracked vehicle maintenance. There was one exception noted where a unit did indeed support, as a garrison requirement, a mechanized unit. Female team members performed well in nine of the eleven units observed. In another unit, of the ten (10) women assigned to the section only one was MOS skilled. In the other, unit no women were assigned to this task even though this is the largest section in the company.

3.2.5.6 Supply Platoon. Again, this element was not organized, equipped or trained, according to its TO&E, deferring to post support requirements. This condition was observed in all cases reported. In several instances, 50% of personnel assigned remained in garrison to continue support of the post, or because of a decision not to take sensitive equipment (computer) to the field.

3.2.5.7 General/Electrical Maintenance. The Mechanical Repair Section deficiencies were similar to those observed in the Automotive Maintence Section reported above. Additionally, in five companies, no ARTEP tasks were performed due to organization shortages of trained personnel or equipment. The Generator Repair Section, however, was a reversal of the usual situation. All tasks were performed well, with sufficient numbers of trained personnel and equipment. The explanation is that tasks performed were those normally performed in garrison and post support missions. Electronic Maintenance Section performance fell between the two sections described above. About half of the observed companies did well. Unsatisfactory performance was due to the same TO&E deficiencies noted above.

3.2.5.8 Service/Recovery Section. Tasks performed were performed satisfactorily in at least half of the units observed. Women performed as well as male counterparts, including wrecker vehicle operation and tire changing tasks. The most often observed discrepancy was that tasks could not be attempted due to lack of ARTEP support (available deadlined equipment).

3.2.5.9 General. Units failed to perform some task or performed them in an unsatisfactory manner due to organizational restructuring, MOS skill deficiency, equipment shortages, lack of ARTEP support, little or no field experience, all of which was reported to be due to the priority given post support mission at the expense of TO&E mission organization and training. Women were usually well integrated into units, especially when chain of command attitude was positive and the first line supervisors were in need of the contribution they could make.

3.2.6 Military Police Units

3.2.6.1 Personnel. The average participating unit was manned at 111% of authorized TO&E. Present for field duty average strength was 81% of assigned personnel. Proportion of women in units was within the test design limits.

3.2.6.2 ARTEP Preparation. The ARTEP plan and scenario was usually closely followed with some schedule modification due to trafficability problems associated with assigned training areas. Cooperation of supporting headquarters was good with one noticeable exception. In this case, a battalion commander contested the value of ARTEP. This required a last minute change of test units and training areas. There was little evidence of concerted effort to conduct pre-ARTEP training, due to post support requirements.

3.2.6.3 Training Status. Generally, MP units organization, equipment and MOS qualified personnel are more closely aligned with TO&E mission requirements than other type units observed in the field. Only four of the test units claimed post support missions interfered with ARTEP despite the fact that all unit perform these garrison requirements..

3.2.6.4 Other. All assigned women went to the field. Women performed satisfactorily assigned tasks and were judged as not to have adversely affected unit performance. This evaluation was unchanged when percentage of women increased from 15% to 35%.

3.3 ARI Staff Visit Trip Reports. Note: Two units with no women assigned were visited. These trip reports have been deleted from consideration.

3.3.1 Sensitivity of ARTEP to measure impact of women on wartime mission performance* Six of the eight reports indicated the ARTEP was

* ARI Comment: These comments should not be construed to mean that the ARI scientists thought the ARTEP was an insensitive measure of wartime mission performance. Based on their observations of units with relatively small percentages of women, and the level of performance of female soldiers, they felt the ARTEP alone without individual performance measures, was not an ideal vehicle for assessing the impact of women on unit performance. In part, this reflected their subjective impressions that an increased fill of women in the units observed would not show an impairment in ARTEP performance. The overriding consideration in using the ARTEP as a measure of performance, was that a standard test be used.

not sensitive for measuring impact of women in wartime mission accomplishment. Two reports were noncommittal. Negative views were based on definitions of wartime missions, lack of leadership measures, no individual tasks which compare male with female, and the relatively short duration (72 hours), and lack of realism and stress.

3.3.2 Extent to which scheduling of events occurred according to scenario, and expanded ARTEP modules were performed and scored. Scenarios and modules were performed according to plan in nearly all cases. Minor deviations were caused by damaged or missing equipment. Major deviations or omissions were caused by lack of cooperation or unwillingness to participate on part of the tested units' higher command.

3.3.3 Attitude of company personnel and local evaluators regarding women soldiers. Women were accepted, but with restraint. They are not viewed as equals. They perform well but present problems like time loss due to sick call, too emotional, physically weaker -- all of which are traditional and culturally shaded opinions. Females did register a disproportionately greater time on sick call than males. The impression of "wait and see," and "what can you do about it," was reported. Again the attitude toward women by company personnel generally reflects the attitude expressed by the higher chain of command. Some exceptions are noted among peers or first-line supervisors.

3.3.4 Performance of MAX-WAC local evaluators, and effectiveness of coordination. Local evaluators were competent and coordination was effective in nearly every instance. When resistence to the ARTEP concept or the idea of women in units was objectionable, coordination was poor and local support and evaluation barely acceptable. This is supported by other data sources reported in this section of the study.

3.3.5 Effectiveness of training for ARTEPS. Training for ARTEPS was hard to judge. In many cases, this was the first opportunity for the unit to get field experience in TO&E mission assignments. In that regard, the training was effective; on the other hand, while training for ARTEPS began with enthusiasm, it was often slowed or discontinued due to higher priority post support missions. In some instances, the realization of the nearly total absence of field/tactical skills overwhelmed the unit and a feeling of futility set in. The prospect for improved training was high for repeated measures units.

3.3.6 What special treatment accomodations were provided women soldiers? Did women fully participate? No special accomodations were provided aside from latrine, bathing and segregated sleeping facilities. Women participated fully in all tasks. Occasionally, some were assigned to traditional roles.

3.3.7 What was your impression of how effective women soldiers were during ARTEP? Hard to evaluate. Women did all assigned tasks within time limts. Usually, women were assisted in high strength tasks, or avoided them (and were allowed to do so). An interesting observation was that women should be compared only with men of equal MOS skill and experience since most are new to the job.

3.3.8 What problems are likely to occur in the future? This question was largely avoided except to note that failure to stabilize evaluators and large personnel turnovers in units would have adverse effect.

3.3.9 Describe the attitudes at installations and the ARTEP events which may be passed on to FORSCOM. How might someone opposed to the continuation of MAX-WAC use the events occurring during this ARTEP to support their position? This question also was avoided. Exceptions are statements that there appeared to be a lack of command emphasis which, if present, would have provided more support and discouraged departures from the test design.

3.4 Hypotheses* Constructed from ARTEP Observations. Of the fifty eight hypotheses examined, forty four were eliminated by (1) combining with other similar statements; (2) because they were not relevant to the MAX-WAC research question; (3) lacked sufficient data to support or not support; or (4) were statements of common knowledge; e.g., "Units do well on those tasks performed frequently..." The remaining fourteen hypotheses are discussed below:

* ARI Comment: These hypotheses were formulated by a MOBDES USAR Colonel, a practicing clinical psychologist, during his two week active duty assignment to the Test Directorate and were based on his reading of the after-action reports and discussions with evaluators. Once formulated, they were given to the team members for comment and further observations.

3.4.1 Hypothesis: Female soldiers assigned to non-traditional MOS positions under conditions of low-fill TO&E tend to be more rapidly assimilated than female soldiers assigned to high-fill organizations or one above its level of TO&E authorization.

Discussion. This is supported by all data sources. In full, or nearly full strength units, women tend to be overlooked and placed in traditional roles. This also occurs when only a small number of women (one to four) are assigned. When the need for personnel is high, as in understrength or overtasked units, women are more readily integrated.

3.4.2 Hypothesis: The recommendations of first line supervisors regarding the duties of female soldiers on the basis of traditional physical statements relating to health status, reflects a markedly conservative supervisory attitude which tends to diminish effective management practices while raising the issue of "double standards" favorable to females.

Discussion. This is supported by all data sources. The average male is unfamiliar with female physiology beyond the level of "folk myths," particularly with complaints associated with the menstrual cycle. Therefore, there is a tendency to misinterpret these complaints and to release female soldiers from duty unnecessarily. This has an adverse affect on utilization of females and unit effectiveness by lowering available work force and morale. This is a symptomatic indicator of the larger obstacle to full utilization of women - - generalized ignorance of female capabilities and limitations fostered by cultural traditions.

3.4.3 Hypothesis: The tendency exists to "protect" female soldiers, as opposed to male counterparts, in certain recognized hazardous situations.

Discussion. This is not well supported. The "protective" male behavior, and exploitation of it by female soldiers, is spotty and is as inconsistent as is the understanding and experience of working with females -- which was very often displayed during the ARTEPS. The absence of definitive policy guidance from higher headquarters allows local commanders to act on their knowledge and experience, thus accounting for the inconsistency.

3.4.4 Hypothesis: Acceptance of female military members by unit CO's and NCO's is positively related to acceptance of military women by their unit male counter parts.

Discussion. This is strongly supported by all data sources. If the chain of command expresses itself positively or negatively toward female soldiers the subordinate elements act out this attitude. It was observed that this was the case from platoon up to post level of command. There was no evidence of disagreement at a lower command level with the attitude expressed at a higher level, and, therefore, there is no information relevant to attempt to reverse or discredit positive or negative statements. It was clear that soldiers do what they are told, or what they believe they have been told.

3.4.5 Hypothesis: Female soldiers function in terms of stamina as favorably or better than male soldiers during field problems requiring short field stays, i.e. three to five days.

Discussion. There was inadequate evidence to support with this statement and little evidence to refute it. This hypothesis is included only because it occupies much of the discussion reported during ARTEPS. The test design did not account for this characteristic, and, therefore, male and female soldiers' differences in this regard were not reported on.

3.4.6 Hypothesis: Leadership, unit training, and experience have greater impact in mission performance than the percentage of females in a unit.

Discussion. This is supported by the data. However, it may be misleading and conclusions should be cautiously drawn because it may be said that leadership, unit training and experience have a greater impact on mission performance than many other factors. It was observed that units which satisfactorily completed mission assignments varied less in training, experience, high morale, leadership, MOS skills, enthusiasm, and operable equipment than they did in proportion of females assigned to the unit.

3.4.7 Hypothesis: Given appropriate training there is no difference between performance of male and that of female soldiers in the construction and maintenance of defensive positions and proper defensive tactics. Examples: perimeter establishment, weapons handling, foxhole preparation, installation and use of tripods as well as traversing and elevating mechanisms for M-60 machine guns.

Discussion. Supported but grossly misleading because of (1) the way the hypothesis is constructed (" Given appropriate training..."), and (2) the recorded observations of tactical tasks indicated that performance was very unsatisfactory. Female soldiers are not given "appropriate training" in tactics or weapons. Therefore, the hypothesis is an assumption supported by assumptions. It should be noted also that while women performed equally well as their male counterparts in tactical operations, neither performed unsatisfactorily.

3.4.8 Hypothesis: Female officers and NCOs are better equipped, especially in the absence of special education programs, to understand

and cope with the variety of physical and psychological complaints and anomalies which affect women.

Discussion. Supported by the data. Widespread comments attested to the fact that male supervisors were inexperienced and unskilled in managing women members of their unit. Many problems associated with female soldiers stem from this institutional ignorance. Appeals for female leaders were more often expressed than appeals for education of male supervisors.

3.4.9 Hypothesis: The successful performance of the vast majority of military tasks requiring team effort is relatively independent of personnel composition, i.e. Whether the team is composed of men, women, or men and women.

Discussion. Supported by all data sources. The test design emphasized unit performance instead of individual performance, changing the proportion of females in the unit. Evaluations indicated no significant degredation or improvement in task performance attributable to changes in the sex composition of teams, other factors being equal (training, experience, attitude etc.). It is interesting to note that even in areas of suspected difficulty -- a sex difference was only marginally noted, e.g. placing litter patients in ambulances.

3.4.10 **Hypothesis:** Male soldiers display a significantly higher tolerance than female soldiers in doing jobs under wet, cold and dirty conditions.

Discussion. Not sufficiently tested, this hypothesis is included only because there was much concern expressed about this subject. One unit suffered higher female than male evacuation due to cold weather conditions. This was insufficient evidence to support or to refute the Hypothesis. Other experiences were not reported.

3.4.11 **Hypothesis:** Unit acceptance of female soldiers is significantly related to willingness to learn the job willingness to respond to a given situation and experience in the task to be performed.

Discussion. Supported by all data sources. In the absence of expressed positive or negative attitudes toward women by the chain of command, women enter units as an unknown quality, and somewhat suspect. When they demonstrate a willingness to learn, try, "join the team," and demonstrate enthusiasim for their work, acceptance is offered, even if at first only tentatively. When demonstrated MOS skill is added, acceptance is nearly immediate at the team member level. Middle level supervisors are slower to respond.

3.4.12 **Hypothesis:** Most military tasks difficult to accomplish using one person are so regardless of gender, i.e. male or female (whether the individual performing the task is male or female).

Discussion. Supported. Strength related tasks were more difficult but not impossible for women to perform. The outstanding example was the inability for women to load litter patients into ambulances or execute heavy lifts and long carries. The field solution was to augment litter teams or mix male and female. In reality, most tasks evaluated were team tasks and were satisfactorily accomplished. Suggestions were recorded that MOS be reviewed, mechanical aids be provided, or male-female team mix be established. Seldom was it observed that women should not perform the task assigned. Some tasks, such as the loading task described, or hand cranking a cold 10KW generator are indeed physically inappropriate for the average female - but they are isolated and not representative. Timeliness of task completion suffered in these instances. It was observed that male-female strength differences could be equalized with training or mechanical aids.

3.4.13 Hypothesis: Field conditions create significantly greater hardships with consequent reduced functioning for female soldiers as opposed to male counterparts.

Discussion. Supported. This has to do nearly exclusively with field sanitation, hygiene and personal privacy. Commanders were hesitant to task male soldiers to prepare female latrines and females were relatively untrained to perform the task. Consequently, most field latrines for females were substandard. In situations where latrines are not prepared, such as breaks during road marches, females experienced greater hardship. This was also observed for situations of clothing changing and bathing. The traditional "bath in a helmet" was not an acceptable solution. It was also observed that this situation was due mostly to the lack of training and innovations; therefore, the support for the hypothesis may be misleading.

3.4.14 Hypothesis: The continuance of pregnant female soldiers (though small in number) on unit strength rolls and in limited duty status creates readiness and morale problems.

Discussion. In the absence of DA Policy, local commanders institute their own policy, which is often uninformed regarding female physiology. The practice has been to relieve from normal duty a pregnant female beyond her third month of term, fearing adverse physical consequences would result from continued full duty. This means no field duty. Pregnancy has been considered a temporary physical disability, and therefore affects readiness only if deployment of the unit occurs during the subject's term. Also, morale problems are reported when the pregnant female soldier is not replaced, and male members must assume the redistributed work load. It is also observed that some male members complain that there is no similar field duty relief for them. In truth, the frequency of this complaint is small but consistent. No uniform policy is available.

4. CONCLUSIONS

4.1 Utilization of Women. Utilization of women is a function of need, e.g. If a unit is understrength or short in specific skills, women will be more rapidly assimilated into the units and used in their MOS rather than in the "traditional" role.

4.2 Protective Attitude. Males tend to be protective of women thereby creating additional workload on the male soldier.

4.3 Degradation of Unit Performance. No degradation of unit performance was noted by the assignment of women soldiers to the units at any level within the test design.

4.4 Acceptance of Females. Acceptance is a function of attitude. The attitude of the chain of command toward women soldiers, whether positive or negative, is reflected by the unit members.

4.5 Basic Training. Women soldiers do not receive adequate basic field training or training in other non MOS skills.

4.6 Attitude of Women. Women soldiers object to being treated in the "traditional" womens' role e.g. being escorted after dark, separated for sleeping purposes, placed in office/clerk positionsrather than in positions for which they are trained.

4.7 Performance of MOS Tasks. Given equal civilian experience and military training women can perform MOS tasks with a proficiency equal to that of men except those which require average male physical strength.

4.8 Women as Team Members. Women are accepted and utilized as team members by first line supervisors if they are MOS qualified or display a willingness to learn.

4.9 Unit Training. Units observed in the test, because of post support requirements were not adequately trained, equipped, or manned to perform the prescribed TOE missions.

DEPARTMENT OF THE ARMY
UNITED STATES ARMY OPERATIONAL TEST AND EVALUATION AGENCY
5600 COLUMBIA PIKE
FALLS CHURCH, VIRGINIA 22041

8 AUG 1977

CSTE-ED

SUBJECT: MAX WAC

Lieutenant General John R. McGiffert
Director of the Army Staff
Office of the Chief of Staff
Washington, D.C. 20310

1. In response to your letter of 16 June 1977, OTEA conducted an independent assessment of the extent to which the MAX WAC test will meet its specified objective. We have also addressed the question of the need for additional evaluations and have included a number of specific recommendations concerning the overall question of women in category II and III units.

2. Although the MAX WAC test results provide much useful information and perceived trends, OTEA's overall conclusion is that the results of the MAX WAC test do not provide a firm basis upon which the Army can make its decision regarding the optimum level of female soldiers in the Army. Rationale for this conclusion and our recommendations are presented in the inclosure.

3. OTEA is prepared to provide support which may assist you as you continue to develop a conclusion to the question of the optimum level of female soldiers in the Army.

1 Incl
as

Julius W. Becton, Jr.
Major General, USA
Commanding

CF:
Commander, US Army Research Institute for Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333
Commander, US Army Training and Doctrine Command, Ft. Monroe, VA 23651
Commander, US Army Forces Command, Ft. McPherson, GA 30083
Commander, US Army Administration Center, Ft. Benjamin Harrison, IN 46216

OTEA
REVIEW AND EVALUATION
OF
MAX WAC STUDY

1. References.

a. Letter, DCSPER, DA, 9 November 1976, subject: "Women Content in Units."

b. Letter, Director of the Army Staff to Commander, OTEA, dated 16 June 1977.

2. Background.

a. For several years the Army has been conducting study efforts intended to address the effective utilization of female soldiers. The most recent formal study in this effort is the MAX WAC Force Development Test and associated study conducted as a result of DCSPER, DA directive to Army Research Institute (Ref 1a). The purpose of the MAX WAC study is to determine what effect variations of female strength in company level units will have on the ability of those units to perform their normal missions. This information is intended to contribute to the Army policy regarding male-female content in each type unit tested. The method of testing chosen to provide the data for the MAX WAC study was to evaluate the performance of a representative sample of units undergoing Army Training and Evaluation Program (ARTEP) exercises. Selected units were tested to determine if the percentage of women in the unit affected unit performance. Ideally, the overall results of MAX WAC could be predictive of optimum mix for the specific type units tested. The MAX WAC study effort is still under way.

b. Recent events caused MAX WAC to be perceived by DA "as a much greater determinant of potential for Army female content than may have been the case when the test was designed" (Ref 1b). As a result, the Director of the Army Staff tasked the Commander, OTEA, to provide an independent review and evaluation of the MAX WAC study effort in the context of recent changes.

3. Purpose and Scope. In a letter from the Director of the Army Staff to the Commander, OTEA, dated 16 June 1977, the following specific objectives were identified for OTEA's review and evaluation effort:

a. To provide an assessment of the extent to which MAX WAC will meet its specified objectives.

b. To determine what remains to be accomplished to establish the optimal female level content in Category II and III units.

c. Based on these first two assessments, recommend any additional tests or evaluations that should be pursued.

4. Approach to the Evaluation. An OTEA task force was organized to examine the concept, design, execution and evaluation process employed in the MAX WAC study. At the time OTEA was assigned its task only one ARTEP remained to be conducted. OTEA's task force observed this ARTEP but did not have sufficient time to conduct independent additional testing of units specifically to evaluate the optimum role and force content for women in the Army. OTEA's assessment would therefore be, in addition to its own observations, to verify the validity of those factors on which MAX WAC results would be based. This would be accomplished by an examination and analysis of the statistical data base collected for the MAX WAC study. It would be augmented by a selected subjective analysis of qualitative data which could be gathered in follow-on visits to units which participated in MAX WAC ARTEPs. As a final step, independent of the structured ARTEP scenarios, the OTEA task force selected for observation an extended free play joint field exercise, BRAVESHIELD, being conducted in the Mojave Desert. This exercise had participation from US Army support elements composed of a high percentage of female personnel. The purpose of this final observer visit was to collect subjective data on durability of women in the field which might confirm or refute the analysis performed on MAX WAC ARTEP units. It was anticipated that it might also provide information which suggested other methods of testing than were available from MAX WAC.

5. Method of Analysis. The methods of analysis on which OTEA's findings and conclusions are based, are discussed in summary below. A more detailed explanation of the various methods, procedures, and the associated results are contained in Tabs A, B, C, and D, and are accordingly referenced in the following subparagraphs.

a. Psychological analysis (TAB A). The following methods were applied to examine the validity of the human factors data collected during the MAX WAC: examination of questionnaires; observation and discussion with participants during an ARTEP; comparison of single and double ARTEP companies based on ARTEP scores; analysis of ARTEP modular scores for missing data, and analysis of ARTEP scoring differences using classical statistical treatment.

b. Statistical analysis of ARTEP ratings (TAB B). The statistical analysis portion of this report analyzes the ratings received by units undergoing the MAX WAC ARTEPs. To analyze this data, a cross-classified design was used, rating double ARTEP units according to the adjectival ratings (outstanding, satisfactory, and unsatisfactory) received in both ARTEPs. These data were counted, sorted and arrayed into 3x3 contingency tables. In this way changes in the ARTEP ratings were observed and analyzed using minimum discrimination information procedures. Appropriate references describing these statistical techniques are annotated in the text at Tab B.

c. Qualitative analysis (TAB C). To determine whether conditions existed in the ARTEP evaluations which could have been confounded to some extent by unidentified conditions or factors present in the tested unit or the conditions of the test, OTEA observers visited a selected unit from each of the five types of units which received a MAX WAC ARTEP. During the course of these visits, the observer team conducted unstructured discussions with personnel from the tested unit, the local command evaluation group, and the exercise controllers. The results of these discussions were important in providing an insight into the attitudes of these personnel and their perceptions of the adequacy of the test, the conditions present which may have influenced the outcome of the test, and their perceptions of the merits of women in their particular type unit. These discussions also contributed to judgmental inferences and findings of this report.

d. Follow-on evaluation (TAB D). As an additional step in examining the utilization of female soldiers in Army units, the OTEA observer team visited a long term joint field exercise where unstructured interviews and observations were made which paralleled the effort conducted on the ARTEP evaluations.

6. Major findings. The results of the OTEA evaluation provided findings of both a statistical and subjective nature. The complete basis of these findings are discussed in the attached annexes.

a. Use of the ARTEP as a test vehicle.

(1) The ARTEP for each of the five types of units evaluated in MAX WAC was developed experimentally. The design and implementation of these ARTEPs was for the specific purpose of MAX WAC evaluations. The usefulness of a previously non-standard measure of unit performance, as a means from which to draw conclusions which are general in nature, is therefore questionable. At best the validity and reliability of these ARTEPs as a measure of unit performance of a type unit is unknown.

(2) Units administered double ARTEPs were brought to TO&E strength level by sudden introduction of female personnel. In many cases the unit was not given sufficient time to stabilize under these new conditions before being subjected to an ARTEP. The result was that in many cases women were too new in the unit to know their jobs or the unit procedures with which they were expected to conform. Conversely, supervisors were limited to their lack of knowledge of the capability of newly assigned individuals. These individuals tended to be newly assigned female personnel introduced to meet the unit's MAX WAC fill requirement.

(3) Many units which were given ARTEPs normally performed a mission in a garrison environment substantially different from their combat mission. The influx of female personnel and its effect were confounded by the task of overcoming a field test scenario for which the unit was not fully prepared.

(4) Increasing the percentage of female fill in a unit was not necessarily accomplished at all levels of grade and chain-of-command structure. Introduction of a certain number of female soldiers in order to meet a fixed percentage of unit strength usually resulted in an over-fill at the lower end of the grade structure and shortages at the upper levels. Such a condition is not representative of the situation that should be expected to exist when women have achieved a proportional distribution throughout the organizational structure.

(5) The administration of a second ARTEP to some units was not conducive to obtaining high unit scores. Units were aware that the second ARTEP was for MAX WAC purposes. Increasing the numbers of females in the unit for the second ARTEP was therefore confounded with varying degrees of attitude change toward acceptance of this challenge.

(6) The use of a relatively short field exercise (approximately three days) allows some personnel to perform temporarily at a higher work output level to meet mission requirements. It is therefore possible, in the case of the MAX WAC ARTEPs, that the results that the unit obtained may not represent what the unit would do if given a long term requirement where all personnel, including women, would be needed to share the workload.

(7) Observations, interviews, and a review of the after-action narratives indicated that there were many variables present, other than the percentage of female fill, which affected the units' ARTEP scores. These included such areas as leadership and command policies. Thus, the ARTEP does not appear to be a direct or positive indicator for measuring the effects of varying female fill.

b. Statistical evaluation of ARTEP ratings.

(1) For double ARTEP units, including control companies, the differences in scores in 11 of 15 companies were statistically significant between ARTEPs. In five of these units, the scores increased, and in six, there was a decrease. (See Tab B, Figure B-4.) This could be indicative of a random process that will provide, in the long term, an equal distribution of unit performance above and below the level of the first ARTEP score. But as a group, certain type units did consistently better than others. This may indicate that an increase in female content is better suited to specific type units rather than a broader classification of units, e.g., Category II or Category III. However, in three of the five types of units receiving two successive ARTEPs, the performance of the control companies was not stable between tests. This variation in ARTEP ratings in the control companies casts doubt on the utility of the ARTEP as a suitable means of satisfying the primary MAX WAC objectives.

(2) There were great variations in the ratings received by single units in the MAX WAC ARTEP exercises. The difference between units, by type, appeared to be greater than the differences between like units with varying female fill. The magnitude of the "unit effect" in single ARTEP companies was approximately 30 times that of the "fill effect" in influencing the ARTEP ratings. However, and although the reason is not evident, there was some indication in the units tested for MAX WAC that units with a higher percentage of female fill performed better than those with a lower fill.

c. Factors affecting female acceptance and performance.

(1) The chain-of-command in units undergoing MAX WAC ARTEPs, particularly at the senior NCO level, was predominantly male. There was a reluctance on the part of male supervisors to deal evenhandedly with males and females alike. Use of female soldiers was, in some cases, a last resort. This appeared to be greatly influenced by a lack of familiarity in dealing with women in a field environment. The case of female NCOs dealing with male subordinates was sufficiently uncommon that no subjective evaluation can be rendered.

(2) Female soldiers were apparently not well trained in field duties, particularly in coping with field conditions and the environment. This was true both of initial military training (BCT and AIT) and in unit training after assignment to operational installations or units. Women interviewed indicated the need for better training in weapons and tactics, and an improved field uniform.

(3) Women generally had a misconception of field duty and somewhat unrealistic expectations of Army life and their jobs based on perceptions held prior to enlistment. This mismatch between expectations and reality can lead to frustration and a lowering of morale.

(4) There are some tasks which involve the use of strength beyond the normal capability of women. These tasks appeared to be few enough in number that, where necessary, women could be assisted or replaced by men to accomplish some jobs. However, comprehensive research may be required to offset the physical disadvantage of women Strategies for this research could include redefinition of jobs, development of job aids, and respecification of equipment design standards. MOS selection standards, for example, might be made gender free so that anyone, regardless of sex, who meets realistic strength and endurance requirements, may be trained for an MOS.

(5) Perhaps the greatest hinderance to utilization of women in military positions is the lack of understanding, and subsequent lack of acceptance of women, based on traditional male-oriented values. This resistance may be strongest at the higher supervisory levels where contact with women is more distant and therefore judgment is not tempered by the reality of contemporary accomplishments. In those units where women are commonplace, their acceptance on individual merit appears to be routine. On the other hand, units of like type where women are not fully integrated may be less receptive to the use of female soldiers, particularly in positions previously within the male domain.

d. Observation on female contributions.

(1) All units surveyed as a part of the OTEA effort, indicated that there are certain duties which females perform better than men. This may be due, in part, to the higher quality female recruit being received. There are, many jobs and MOSs ideally suited to women, or where women perform equally as well as men.

(2) Unit commanders were quick to indicate that, generally, women were less of a disciplinary problem than men, and therefore, more reliable. Reliability was, in fact, often mentioned as a strong point irrespective of discipline.

(3) The female contribution to the unit appeared to be looked on most favorably by their male peers. The longer the exposure to female partnership, the more routinely the women seemed to be accepted.

(4) In most units visited, the commanders expressed skepticism on the ability of women to endure long term stress. This perception appeared to be based on preconceived, male-oriented values, rather than experience. However, the OTEA visit to exercise BRAVESHIELD tended to dispel the notion that women could not endure the hardships of the field environment for an extended period (see paragraph 6e).

(5) A uniform concern of all commanders interviewed during the conduct of the OTEA evaluation, was that of pregnancy among female soldiers. While there were varying figures posited by each commander as to loss rate and decrease in unit mission effectiveness due to pregnancy, it was evident that there is considerable doubt at the unit level on how to deal with this problem. The OTEA team found no evidence of a command effort to discourage, prevent, or terminate pregnancies in the units. Although identified as their most serious problem, there was reluctance by unit commanders to deal with the subject in the absence of any higher level policy guidance.

e. Observations of long term stress situation.

(1) Mostly through lack of adequate training in basic soldierly field techniques, women appeared to require more time to adapt, initially, to field duty. Those women observed during Exercise BRAVESHIELD, however become as well acclimated to the field and the severe desert environment as the male soldiers. There were no differences noted in the performance of women as compared with men. There were a number of problems in the field situation, however, which were a result of inadequate unit planning for some female-peculiar requirements. These included the need for sufficient separate latrine and shower facilities and the requirement for a certain minimum degree of privacy.

(2) Females appeared to withstand the extreme heat as well as their male counterparts.

(3) Women performed their duties, in the opinion of superiors and peers alike, in a manner equal to male counterparts.

(4) There were no serious social or disciplinary problems observed as a result of the presence of female soldiers.

(5) There appeared to be a lack of realization among the women that their duties, i.e., combat service support functions, were part of a combat scenario which in time of war could put them in a situation of great peril. In discussing this matter with those women interviewed, there was an obvious lack of realization of the relationship of their duties to a combat situation.

(6) The long term free play exercise showed promise as a vehicle to evaluate women in the field because of the stabilized and relatively realistic conditions. Most of the data which could be gathered under these conditions, without overburdening the units with a large group of evaluators, or much non-exercise related work, would necessarily be subjective in nature. There are, therefore, important methodological considerations to such a proposal. These are discussed in detail in Tab D.

7. Conclusions and Recommendations.

a. Conclusions.

(1) The MAX WAC study does not provide an empirical basis to objectively support establishment of an upper bound on potential female content of military units. However, the OTEA effort subjectively determined that in those types of units examined, there were no apparent serious problems detectable at about the 20 percent fill level, notwithstanding specific detailed problems in individual MOSs.

(2) The percentage of female fill in a unit should be addressed in terms of the percentage of female fill within each MOS of that unit. This was not done in MAX WAC, and therefore, any conclusions on optimum unit mix, may be unreliable.

b. Recommendations for determining an optimum female level content in Category II and III units.

(1) The Army should pursue with vigor the evaluation of the entire MOS structure being undertaken by the Admin Center to determine specific strength and skill requirements in individual MOSs. This effort should provide a basis for determination of the maximum/minimum male-female mix in unit TO&Es by MOS.

(2) As a corollary to the MOS study, the role of women in unit self-defense needs to be clearly defined to determine if there is a limitation imposed by females in Category II and III units.

c. As a long-term effort beyond the MAX WAC studies, it is recommended that such evaluations concentrate on the systematic observation of extended field exercises which will better exemplify the performance of women in relatively stabilized and realistic combat scenarios and where detailed MOS-related contributions will be more evident. In addition, previous studies should be examined, and interviews conducted with key personnel in units containing female soldiers.

d. Although not identified as specific objectives for the OTEA review and evaluation, several general recommendations on female soldiers evolved from this effort.

(1) In orienting leaders and soldiers in the role of women in the Army and techniques for effective leadership of female soldiers, high priority should be given to establishing training at the entry level, branch and service schools, NCOES, and in mobile training teams.

(2) Women should be accepted as soldiers and not as females. An immediate step forward in this issue would be the integration of Basic Combat Training so that all soldiers are similarly trained in entry level soldierly skills.

(3) The Army should establish and promulgate guidance to the field in handling pregnancy problems, fraternization, and billeting.

(4) Based on numerous complaints made by female soldiers, the design and quality of material in female uniforms needs to be brought to the level of male clothing if females are to be expected to endure similar field conditions.

TAB A

PSYCHOLOGICAL ANALYSIS

1. **Discussion.**

a. The MAX WAC study used a company's ARTEP score as the measure of effectiveness for unit performance. To obtain an ARTEP score for a unit, a three to four day field exercise was used with a standard scenario for a type company. A team of independent evaluators then scored selected tasks, called modules, on a three point scale:

1 - the task was not completed

2 - the task was completed in an average manner

3 - the task was completed in an above average manner.

A company's ARTEP score was the average of its module scores for those modules which were scored. No attempt was made to weight the modules in deriving the ARTEP score; a company's score was not adjusted for the number of modules which were used; and no weighting was made for the different number of modules composing each type of ARTEP.

b. The MAX WAC study observed five types of combat service support units: maintenance, medical, military police, signal, and transportation. ARTEPs were developed for the MAX WAC evaluation for these types of units. Consequently, the reliability and validity of those ARTEPs were unknown. Five companies of each type unit were administered one ARTEP each, during the period October 1976 to June 1977. These are referred to as single ARTEP units. Additionally, three companies of each type were administered an ARTEP twice, once during the period October 1976 to December 1976 and once during the period January 1977 to June 1977. These are referred to as double ARTEP units.

c. The double ARTEP units constituted the experimental and the control units for the MAX WAC study. For the first ARTEP administration, one company of each type was filled with 0% women and tested, one company was filled with 15% women and tested, and one company was tested at whatever its female fill percentage happened to be. The latter was a control company. Prior to the second ARTEP administration, those double ARTEP units with no women were brought to 15% women, those with 15% women were raised to 35% women, and the control companies were to remain as they had been. Changes in fill level were to be accomplished no later than 60 days prior to an ARTEP administration to allow perturbations from these changes to smooth out. Officers and noncommissioned personnel were stabilized during the test. Roughly six months was to elapse between tests.

d. Control units were supposed to be maintained at their original fill level between the first and second ARTEPs. The purpose of these units was to provide an indication of how ARTEP scores might change between administrations when the percentage of women was undisturbed. This was needed because the Army had no experience with the ARTEPs' reliability since these ARTEPs were developed as part of this research effort.

e. Special purpose questionnaires were administered to officers, noncommissioned officers, and enlisted personnel after each ARTEP to tap aspects of the test situation and social milieu not addressed by the ARTEP measure of effectiveness itself.

f. Whatever results from statistical analysis of the ARTEP data, the generalizability of the outcome is severely restricted. Reasons for this restriction are discussed below in terms of uncontrolled sampling, atypicality of experimental companies, uncontrolled variables, and missing data.

2. Design Limitations in Test Execution .

a. From its inception, the MAX WAC study was never classically pure in a design sense in that the sample of 40 units used was neither a random nor a representative sample of similar Army units, either in or outside CONUS. This is in part due to FORSCOM being the agency which designated the units to participate in the study.

b. A second design question is to ask the extent to which the experimental and control companies initially compared with units of their type. If one assumes that single ARTEP companies while not a representative sample, are not altogether a bad sample, then one can use the ARTEP results for the 25 single ARTEP companies as a standard by which to judge the first ARTEPs of the double ARTEP companies. By this criterion, the double ARTEP companies were atypical and ranged from extremely poor to excellent. The mean and standard deviation of the five ARTEP scores for each type of single ARTEP company were calculated as shown in Table A-1. Each double ARTEP company's first ARTEP score was then scaled by the following transform:

$$Z = \frac{\text{(Company Score) - (Company-type Mean Score)}}{\text{Company-type Standard Deviation}}$$

By this measure, the 15 double ARTEP companies ranged from eight standard deviations below the mean, to five standard deviations above the mean, as shown in Table A-2. Variations this large make the double ARTEP sample suspect in its ability to provide results which would be meaningful for units of the same type.

Table A-1. Means and Standard Deviations for the ARTEP Administration of the Single ARTEP Companies.

	Mean	Standard Deviation
Signal	1.91	0.12
MP	1.76	0.31
Medical	2.09	0.09
Trans	2.36	0.16
Maint	2.25	0.07

Table A-2. Z Score Transforms for First ARTEP Administration of the Double ARTEP companies

	0-15 Group	15-35 Group	Control Group
Signal	.50	1.33	1.83
MP	.65	.45	1.13
Medical	2.00	-.89	4.67
Trans	2.00	-.81	.56
Maint	-2.71	-8.14	5.14

3. Uncontrolled Variables in Test Execution. A number of uncontrolled variables are associated with test execution. These occur at the Army level, the MAX WAC study level, the installation level, and the unit level. It should be noted that this breakout is somewhat arbitrary, and serves only as a way of organizing these variables.

a. The Army Level.

(1) The two main limitations to the results of the Army Level are that few women currently have entered the ranks of noncommissioned officers, and that current male noncommissioned officers are largely inexperienced in dealing with female soldiers. Women are now entering more MOSs than ever before, but they have not been in their MOSs long enough to have become NCOs. Consequently, what impact women serving in leadership roles in the enlisted ranks will have, remains to be seen in the Army generally; and specifically, in the present study, it was lacking altogether.

(2) Second, many male NCOs are unsure of how they should deal with female soldiers and are sometimes overly lenient with them in task accomplishment. Consequently, an additional load is sometimes imposed on the male soldiers to accomplish the unit's mission, but at the same time, women are denied the opportunity to demonstrate their competence and inadequacies. Just as this is a problem for the Army generally, so too

it was a problem for the MAX WAC study, particularly because the ARTEP scores are derived from module accomplishment, but do not in and of themselves indicate who in the unit was responsible for the success or failure of the task. Presumably, there are NCOs in the Army who are overly demanding of female soldiers, but examination of the enlisted personnel questionnaire comments did not surface any instance of this.

b. The Study Level.

(1) Five limitations may be noted at the Study Level. They all introduce unknown variability unevenly applied to the ARTEP measures of effectiveness (MOE). The first is variation in the workload under which units operated between their first and second ARTEPs. Sometimes a unit took one ARTEP as an integrated part of a full-scale division exercise, and the second ARTEP, as a separate company level exercise. Consequently, any effect of percentage of females in the unit was obscured by differences in the degree of tasking of the unit from one ARTEP to the next.

(2) The garrison mission for a unit was sometimes different from its field mission. The ARTEP modules were derived for the field mission. The consequence is that some soldiers did not exercise in the ARTEP the skills they ordinarily used during the rest of the year. It may be argued that a unit's field mission is its combat mission, and that unit commanders are responsible for maintaining the unit's combat readiness. Whatever the merits of the argument, the point is that some units apparently did not train extensively to prepare for the ARTEPs, so that the effect of women in a mock combat situation was not tested under equal levels of training preparedness.

(3) Tasks during the ARTEPs were occasionally done out of scenario sequence and were deliberately assigned to women for execution. This is contrary to normal practice and policy and somewhat alters in unknown ways the validity of the ARTEP score as a measure of effectiveness.

(4) One double ARTEP unit was administered its second ARTEP two months after its first, whereas the remaining double ARTEP companies had from four to seven months between ARTEPs. The quick succession between ARTEPs for this unit appears to have negatively influenced the installation level's command policy and attitude and the motivation of the unit to do well.

(5) The final limitation at the study level is that another control unit experienced a 14% drop in its female complement between the first and second ARTEP. Since the purpose of the control units was to gain some insight into the direction and magnitude of change in ARTEP scores for repeated measurements while percentage of female fill was undistrubed, the 14% drop was detrimental to the validity of the study.

c. The Installation Level. The major limitation to the MAX WAC study at the installation level was the occurrence of instances of negative command policy, attitude, and willingness to support the program. This type of attitude appears to have then permeated throughout the installation and probably had an effect on unit performance. For example, women were attached to units rather than assigned, so that normal processes of incorporating new personnel into a company were deflected. The consequence was to increase the artificiality of the MAX WAC study. Further examples are that women occasionally were assigned/attached to units only 30, and in one instance, only 15 days prior to an ARTEP. In the latter case, resistance was so strong that command action had to be taken to meet the experimental requirement. Also, inexperienced local evaluators were sometimes used to oversee the ARTEP operation rather than providing more experienced people. Therefore, ARTEPs may have been conducted under less than optimal circumstances. Also, the MAX WAC independent evaluators had to rely on local evaluators' opinions as to whether a task was accomplished in an outstanding manner. Switchover from experienced to inexperienced personnel renders these judgments somewhat questionable.

d. The Unit Level.

(1) A number of limitations to the MAX WAC Study are notable at the unit level. In some cases, unit leadership and/or organization were poor. In other cases, units lacked prior field training for as much as a year prior to the ARTEP. Some units had the attitude that the second administration of the ARTEP was not "for real" because no one's career was riding on the results.

(2) It is unclear in the study if units utilized women in a consistent fashion. For example, it is important to know whether female soldiers were used in their MOSs during the ARTEPs or not, whether they had practiced their MOS skills and were current or not, and whether they were treated differently on these from male soldiers. Examination of the enlisted personnel collateral questionnaire showed that some companies were asked whether personnel had practiced their MOS skills in the last 60 days, but other companies were not. Consequently, it may not be possible to determine a firm answer to this issue from the questionnaire data.

(3) Another example of variation in the utilization of women was their employment by companies in perimeter defense. Some companies assigned women as an integrated member of a foxhole team; other companies assigned women in pairs to foxholes; and others used women on the perimeter during the day but not at night.

(4) Aside from the prior limitations at the unit level, the worst limitation in the MAX WAC study from an experimental point of view is that it appears some units obtained and practiced the specific ARTEP scenarios they were to be tested under prior to the ARTEPs. This is contrary to

normal usage and policy for running an experiment and severely damages the validity of the ARTEP as a measure of effectiveness because a unit which had practiced the scenario may be expected to do so spuriously well in the field exercises. Fortunately the number of instances is small, but this compounds the already difficult problem of interpreting the results of the MAX WAC study.

4. Missing Data Limitations.

a. The percentage of ARTEP modules which were not scored during the 55 ARTEPs is a procedural limitation of the MAX WAC study because the ARTEP scores within a company type are based on observations of different modules. That is, some ARTEP scores are based on 10% missing data and others on 20% missing data. Table A-3 shows the average percentages of missing data and their ranges by company type for the 55 ARTEPs. Each range is across 11 ARTEPs. For example, the range of percentage of missing data for the maintenance companies ran from 5% to 40%. This means that only 18 out of the 40 modules which are used to derive the ARTEP scores are usable for comparison purposes across all the maintenance company ARTEPs if one wishes to do a module by module comparison. For the double ARTEP maintenance companies which had repeated ARTEP measurements, only 23 out of the 40 modules have complete data for a module by module comparison (58%). For the other type of double ARTEP companies, the percentages of modules which have complete data are as follows: Signal - 54%, Military Police - 87%, Medical - 71%, and Transportation - 100%. Variations this large in ARTEP score composition makes the validity of the ARTEP scores as a measure of effectiveness suspect for comparisons within and across company types.

Table A-3. Average Percentages and Ranges of Missing Data for 55 ARTEPs

	Average Percent Missing Data	Range of Percent Missing Data
Signal	18.36	11 - 30
Military Police	2.91	0 - 6
Medical	9.27	0 - 23
Transportation	0.	0 - 0
Maintenance	21.09	5 - 40

b. ARTEP Inappropriate.

(1) Supplementing the restrictions noted previously, it is probably the case that the manner in which the ARTEPs were conducted is inappropriate for assessing present and future impact of women in combat service

support units. The first reason is that the ARTEP as a three or four day field exercise is too short to elicit long-term problems of adjustment both in terms of peer acceptance and in terms of job performance. For a three day exercise, male company personnel can too easily ignore the female complement and take over whatever deficiencies the women may evidence.

(2) No female NCOs were used during these ARTEPs. Consequently, the MAX WAC experiment cannot delineate whatever problems might emerge when females occupy key leadership positions. ARTEPs are sensitive to the performance of key personnel.

(3) The females in the Army now constitute a highly selected group of soldiers. By and large they are Category I and II, whereas entering males are more typically Category III. Since it is unclear whether this relatively high standard can be maintained with a larger influx of female soldiers, the results of the ARTEPs employing women who are essentially pioneers, may well be inappropriate for non-pioneer females of the future.

5. Problems for Parametric Statistical Analysis.

a. One approach which might be used to address the MAX WAC objectives is parametric statistical treatment of the ARTEP data at the modular level. As noted previously, modular scores are overall scores for groupings of similar tasks which were scored during an ARTEP. A company's ARTEP score is the mean of its modular scores. The amount of missing data noted previously poses a problem for analysis of ARTEP scores using such parametric statistical treatment. In contrast to this approach, Tab B presents OTEA's statistical analysis which will be performed on the individual ARTEP scores. This analysis uses each single ARTEP task on a line-by-line basis and not the modular technique as in the parametric approach described herein. On this account, the sample sizes used in the statistical analysis will be larger, thereby increasing its sensitivity to changes in ARTEP scores.

b. Following is an overview of what can be learned using the parametric approach.

(1) For any given company which took the ARTEP twice, one can examine changes in the ARTEP scores by averaging the difference between module scores measured on each occasion. Consequently, any module which was scored only once would be discarded. Analysis could then proceed on the basis of the fifteen average difference scores for the double ARTEP companies, but ARTEP difference scores for companies of the same type, say maintenance companies, would be based on somewhat different modules. For example, company A may have been assessed twice only on modules 1, 2, 3, and 4, while company B may have been assessed twice only on modules 1, 3, and 4. Therefore, one problem is that if all of the data

available for each company is used, comparison of companies of the same type will be unfair because different modules were used to generate the difference scores, and comparisons of groups of companies of differing types will similarly be affected. A solution for this could be to use only those modules on which complete data is available for all companies of the same type in computing average difference scores, but this would result in the use of only a limited portion of the data (from 54% to 100% depending on company type).

(2) If this approach were pursued, two analyses of variance could be run on the ARTEP difference scores for the double ARTEP companies. The first, shown in Table A-4, would use paired modules within each company to generate the ARTEP difference scores. The second shown in Table A-5, would use paired modules across company type to generate the ARTEP difference scores. Both analyses would test the null hypothesis of no differences between the 0% to 15% group, the 15% to 35% group, and the control group from the first to the second ARTEP administration. Both analyses would show, for $\alpha = 0.10$, no discernible effect between the three groups.

Table A-4. Analysis of Variance Based on Paired Modules Within Each Company

Source	SS	df	MS	F	
Groups	.252	2	.126	2.500	Not Significant
Residual	.605	12	.050		

Table A-5. Analysis of Variance Based on Paired Modules Across Company Type

Source	SS	df	MS	F	
Groups	.304	2	.152	2.60	Not Significant
Residual	.702	12	.059		

(3) Given a finding of no difference between groups, it is legitimate to ask how valid the finding is and what the finding says about employment of women in the Army. It should be noted that the analysis does not address whether companies of the same type changed from one ARTEP to the next, but whether one group of companies of different types, on average, changed more than another group. Consequently, one is unable to say whether given an effect due to women had been found, the effect differed by company type. Beyond this, the validity of a finding which would result from an analysis of this type is doubtful for at least two reasons. The first is that no control group failed its intended purpose since one company experienced a 14% drop in female fill between ARTEPs, another company had only two months between ARTEPs, and another company had negative indicators on workload, strength, and higher command policy on the second ARTEP, but not on the first. Consequently, without an adequate

control group, the meaningfulness of any statistically significant difference in the other two groups is lacking. The second reason for doubting the validity of an analysis of this sort is that this and subsequent discussions show the existence of a number of potentially confounding factors.

(4) Taken collectively, the problems associated with the amount of missing data, the instability of the control groups, and the number of confounding variables, make questionable the utility of this type of analytical procedure to assess the MAX WAC data.

6. Non-ARTEP Findings.

a. Four additional findings are noteworthy from the MAX WAC exercise. The first is that some female soldiers had unrealistic expectations about what Army life would be like and what their jobs would be like. They had images of a light vehicle driver being someone who drove a sedan, and were dismayed to learn the Army considered a two and one-half ton truck a light vehicle. The disparity between expectation and reality undoubtedly influences reenlistments as well as attitudes.

b. The second finding is that female soldiers received limited training in weapons usage and tactics, both in BET and AIT. Female soldiers were observed picking up their weapons during an attack and then not knowing where to go. Others were assigned to operate an M-60 machinegun, but were not qualified to do so.

c. NCOs and officers are by and large inexperienced in utilizing female soldiers. NCOs are particularly subject to allowing female soldiers to get by with behavior which they would find unpermissible for a male, in part due to role conflict between being a male and being an NCO. They also assign men and women to do a job, but allow the women to stand by while the men work.

d. The fourth finding is that pregnancy was a universal concern of the unit commanders interviewed as part of this evaluation, but none had taken command action either in easing access to contraceptives or in exercising moral suasion to prevent unwed pregnancies. Clearly, high level Army guidance is required to assist local commanders in this matter.

TAB B

STATISTICAL ANALYSIS

1. Discussion. The key to determining what effect variations of female strength in company level units had on the ability of those units to perform their mission lies in measuring those changes in the ARTEP scores that can be attributed to changes in female strength. To permit such measurements to be made, certain underlying conditions have to be satisfied. These are discussed in the following paragraphs.

a. ARTEP scores should actually reflect the capability of a unit to perform its mission. If this condition is not satisfied, then the ARTEP is not a suitable device for satisfying the test objectives.

b. ARTEP test conditions should be sufficiently controlled so that any changes in ARTEP scores are due to increases in the proportion of the women in the test units and not due to the influence of other experimental variables. Some of these variables are listed below. They apply specifically to those units which received more than one ARTEP.

(1) Leadership. The same leaders should command during both ARTEPs so that the quality of the leadership is constant for both ARTEPs.

(2) Evaluators. The same group of evaluators should score both tests so that there is consistency in rendering evaluations across both tests.

(3) Scenario. The scenario for the two tests should be the same in order to permit consistency in leadership and evaluation.

c. In most experimental situations more than one factor (variable) affects the outcome of the experiment. Through statistical design it is often possible to minimize or even eliminate these extraneous influences by "blocking." In this way each block, such as the unit undergoing an ARTEP, acts as its own control. For companies receiving two ARTEPs, it is assumed that any extraneous factors will affect both sets of scores in exactly the same way. When the two sets of scores are subtracted, those extraneous factors are removed. For example, poor leadership will affect both scores in a negative direction. However, if poor leadership is exercised at the same level in both paired ARTEPs, subtracting the scores will remove the effect of poor leadership since it affected both sets of scores in the same way. Such designs are often called paired designs.

d. It has previously been pointed out that other extraneous, uncontrolled factors were at play in the MAX WAC test. This analysis will also support this notion. For analytical purposes, however, this analysis will be conducted as though the ARTEP test conditions were sufficiently controlled so that any changes in ARTEP scores are due to increases in the proportion of women in the test units and not due to the influence of these extraneous variables. However, the impact of these extraneous variables on the results of the statistical analysis will be considered in the evaluation of all those factors affecting the performance of the MAX WAC units.

e. To analyze the present set of data, a cross-classified design is used. The ARTEP scores are cross-classified according to the number of unsatisfactory, satisfactory, and outstanding scores received in the two ARTEPs.

2. Approach to Analysis.

a. Double ARTEP Companies. Adjectival ratings (outstanding, satisfactory, unsatisfactory) were scored on both the first and second ARTEPs, and were counted, sorted, and arrayed into 3x3 contingency tables. In this way any changes in the ARTEP scores are more easily captured and analyzed. Further details concerning the cross-classification of ARTEP scores will be presented along with the display and analysis of data.

b. The data stemming from the test is count data (discrete data). It is arrayed initially in 3x3 contingency tables, and later in 3x3x3 contingency tables. The principle of minimum discrimination information estimation is used.[1] To test for marginal homogeneity in the 3x3 contingency tables, the procedure calls for comparing cell "estimates" with the actual observed data in each cell of the contingency table. The "estimated" values are those that would be expected if the null hypothesis is true, i.e., increases in the proportion of women in ARTEP units does not impair performance. In this kind of problem, restraints are determined by the hypothesis being tested. The basic point of concern is whether the "observed" values and "estimated" values are consistent with the hypothesis of interest. The information number is expressed in the form $2I(x_a^*:x)$ where x_a^*, as a vector, represents the estimated or predicted values and likewise x represents the actual observed cell entries taken from the ARTEP rating forms. Basically, $2I(x_a^*:x)$ compares an estimated table with a predicted table.[2] Small values support the null hypothesis. Larger values indicate that the null hypothesis should be rejected. The mathematical details are contained in the reference in Footnote 1. Interpretation of the minimum discrimination information statistic, $2I(x_a^*:x)$, used in this report, will be somewhat abbreviated for clarity.

---

[1] Kullback, Solomon, The Information In Contingency Tables, Final Technical Report, September 1974, USAARO Grant Number DAHCO 4-74-G-0164.

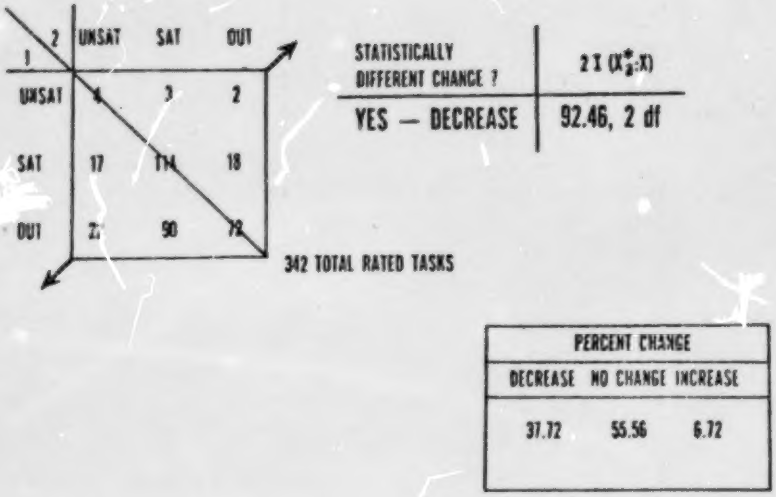
[2] The expression $2I(x_a^*:x)$ will be used for the paired-design case. For the unpaired analysis the expression $2I(x:x_a^*)$ will be employed.

3. Analysis.

a. Unit designations are not shown in order to protect the identity of the company size unit taking the ARTEP. This omission does not affect the findings in any way.

b. Double ARTEP Companies. Three actual cases will be studied in detail to: (1) illustrate the cross-classification procedure, and (2) provide a basis for addressing the principal study objective. Summaries of performance data for the remaining 12 companies will then be made. Findings based upon an analysis of these data will require an analysis of data aggregated by group classification (i.e., control group, 15%-fill group, and 35%-fill group). For example, do the 15%-fill and 35%-fill companies differ from the control group? Finally, an analysis of the five control companies will be made, followed by a corresponding analysis of the 25 individual companies which participated in single ARTEP evaluations.

(1) Medical Company (Control Group). Referring to the 3x3 contingency table, Figure B-1, the following points merit attention.



1
2
UNSAT
SAT
OUT
UNSAT
4
3
2
SAT
17
114
18
OUT
22
90
72
342 TOTAL RATED TASKS
STATISTICALLY DIFFERENT CHANCE ?
2 I (X*2:X)
YES — DECREASE
92.46, 2 df
PERCENT CHANGE
DECREASE
NO CHANGE
INCREASE
37.72
55.56
6.72

Figure B-1. Medical Company (Control Group).

(a) The number 1 represents the categories of ratings for the first ARTEP. Vertically, beneath the number 1 are the three categories of ratings; unsatisfactory, satisfactory and outstanding. The numbers in each row of the contingency table total to the number of these ratings awarded in the first ARTEP. For example there were nine (4+3+2) unsatisfactory ratings in the first ARTEP. Likewise, the same three kinds of ratings are shown horizontally after the number 2 for the second ARTEP. The numbers in each column total to the number of these ratings awarded in the second ARTEP. For example, there were 43 (4+17+22) unsatisfactory ratings in the second ARTEP. Clearly, unit performance fell off in the second ARTEP, as indicated by the increase in the number of unsatisfactory ratings. Since, the same number of line items (tasks) were scored on the two tests, this increase in unsatisfactory scores was made at the expense of other, higher ratings.

(b) A total of 342 tasks were rated for each ARTEP.

(c) Numbers along the diagonal represent ratings for those tasks which remained unchanged. For example there were 114 satisfactory scores on the first ARTEP which were also scored as satisfactory on the second ARTEP. It is important to note that these scores were for the same 114 tasks.

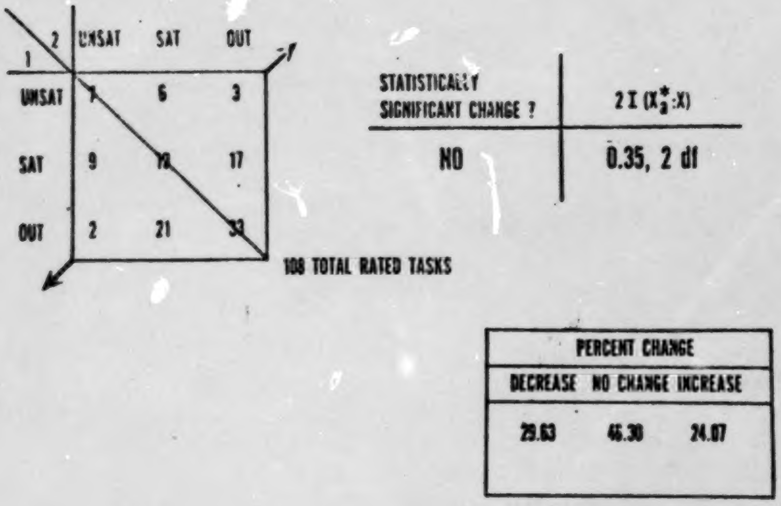
(d) A total of 90 outstanding scores received on the first ARTEP, were changed to satisfactory on the second ARTEP. Seventeen satisfactory scores from the first ARTEP were scored unsatisfactory on the second ARTEP. Again, these changes were for the same line items (tasks). Accordingly, numbers in the lower triangle represent decreases in performance.

(e) Numbers in the upper triangle represent improvement. For instance, 18 satisfactories were raised to outstandings and 3 unsatisfactories were changed to satisfactory, indicating improvement.

(f) The percentages in the lower left hand box indicate the magnitude of these changes. It is noticed that 55.56% of the task ratings remained unchanged across the two ARTEPs. This percent is obtained by taking the total of the numbers along the diagonal and dividing it by 342.

(g) The minimum discrimination information (MDIS) statistic, $2I(x^*:x_a) = 92.46$ with two degrees of freedom, is highly statistically significant. The critical level for the MDIS, which is distributed asymptotically as a Chi-Square random variable ($\alpha = 0.05$), is 5.99. The magnitude of this statistic indicates that a major change in rating scores has taken place and is not due to chance variation. On balance one could conclude that company performance was very different between the two tests and that it decreased considerably during the second ARTEP.

(2) Transportation Company (Control Group). In analyzing the 3x3 table, Figure B-2, the following important points can be observed.



1 \ 2	UNSAT	SAT	OUT
UNSAT	7	6	3
SAT	9	10	17
OUT	2	21	33

108 TOTAL RATED TASKS

STATISTICALLY SIGNIFICANT CHANGE ?	$2I(X_a^*:X)$
NO	0.35, 2 df

PERCENT CHANGE		
DECREASE	NO CHANGE	INCREASE
29.63	46.30	24.07

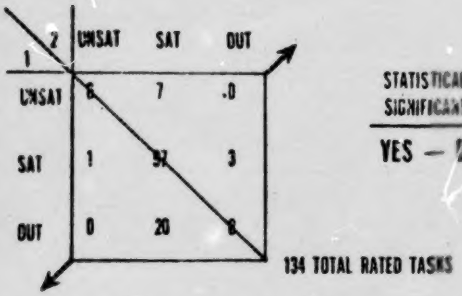
Figure B-2. Transportation Company (Control Group).

(a) There were 108 tasks rated for both tests. Nearly one-half of the ratings remained unchanged.

(b) 21 outstandings in the first ARTEP were lowered to ratings of satisfactory in the second ARTEP, while 17 satisfactories were raised to outstanding.

(c) $2I(x_a^*:x) = 0.35$ with 2 df. This indicates that while some categorical ratings were changed negatively, others increased positively and on balance unit performance did not appreciably change. For example, the 21 outstanding scores on the first test that changed to satisfactory on the second test were offset by the 17 satisfactory scores on the first test which were subsequently raised to outstanding on the second one.

(3) Signal Company (15 - 35% Fill). Referring to the 3x3 contingency table, Figure B-3, the following points are noted.


2
1
UNSAT
SAT
OUT
UNSAT
6
7
0
SAT
1
97
3
OUT
0
20
0
134 TOTAL RATED TASKS

STATISTICALLY SIGNIFICANT CHANGE ?	$2I(x_a^*:x)$
YES — DECREASE	21.25, 2 df

PERCENT CHANGE		
DECREASE	NO CHANGE	INCREASE
15.67	76.87	7.46

Figure B-3. Signal Company (15-35% fill).

(a) 134 tasks were rated on both ARTEPs. Nearly 77% of the tasks were graded the same on both tests. However, evidence indicates that performance declined over the two testing periods.

(b) There were a total of 21 line items (lower triangle) awarded a lower classification in the second tests and only 10 line items showed an improvement in the second test. The value $2I(x_a^*:x) = 21.25$ with 2 df, indicates an important net change in ARTEP scores. Roughly, there were twice as many declines as improvements in task performance and this difference is statistically significant, notwithstanding the fact that 77% of the scores were unchanged. Overall, it can be concluded that this company's performance, over the two ARTEPs, was very stable for the most part, but with a slight decrease in performance during the second test.

(4) Figure B-4 is a tabular summary of the statistics for the double ARTEP companies. It is worthwhile to note that the $2I(x_a^*:x)$ values for the first 4 companies are not statistically significant, while for the remaining 11 companies these values are statistically significant.

UNIT	% FILL LEVEL		$2I(X^*_a : X)$	PERCENT CHANGE			NET CHANGE % (±)
	1st ARTEP	2nd ARTEP		–	0	+	
TC	0 CONTROL	0 CONTROL	0.35	29.63	46.30	24.07	– 5.56
TC	0	15	1.46	23.85	56.88	19.27	– 4.58
MP	8.3 CONTROL	11.7 CONTROL	2.36	26.45	56.20	17.36	– 9.09
MED	0	15	2.48	21.64	52.46	25.90	+ 4.26
MED	15	35	7.57	22.67	52.80	24.53	+ 1.86
MAINT	0	15	8.28	13.95	58.72	27.33	+ 13.38
SIG	0	15	10.56	15.74	80.56	3.70	– 12.04
MAINT	12.9 CONTROL	9.8 CONTROL	12.72	16.67	53.33	30.00	+ 13.33
SIG	24.1 CONTROL	10.2 CONTROL	13.60	28.86	61.86	9.28	– 19.58
MP	0	15	18.74	23.08	57.26	19.66	– 3.42
MP	15	35	21.25	22.69	57.98	19.33	– 3.36
SIG	15	35	21.25	15.67	76.87	7.46	– 8.21
TC	15	35	27.33	12.04	54.63	33.33	+ 21.29
MAINT	15	35	57.67	7.45	45.96	46.58	+ 39.13
MED	24.5 CONTROL	21.6 CONTROL	92.46	37.72	55.56	6.72	– 31.00

Figure B-4. Summary table for double ARTEP units.

c. Figure B-5 summarizes the cross-classification for the 15 double ARTEP companies. It also poses two points of view which challenge each other.

TYPE UNIT	NUMBER CHANGES	
	DECREASE	INCREASE
MAINT	0	3
MED	1	1
MP	2	0
TRANS	0	1
SIG	3	0

Figure B-5. Shifts in assignment scores by type units.

(1) On the one hand there were five increases and six decreases in unit performance that were statistically significant, or about as many increases as decreases. This could be indicative of a random process that in the long run will yield as many ups as downs in unit performance.

(2) On the other hand, maintenance companies scored improvements while MP companies fell off in unit performance. Therefore it might be said that perhaps women in the Army do better in maintenance units than in MP units. However, it should be pointed out that both comparisons also include the control companies. To explore this notion further, the ARTEP scores within the five selected types of military organizations were analyzed.

(3) As stated in paragraph 1d, both points of view expressed in paragraphs 2c(1) and (2) above are affected by extraneous uncontrolled variables. Although their effect is not noted in the statistical analysis, their impact, if it can be determined, will be considered in the overall evaluation.

d. Consistency of ARTEP Scores Within Type Military Units.

(1) Figure B-6 depicts the medical control unit 3x3 contingency table together with the 3x3 tables for the 0-15% and 15-35% fill medical units.

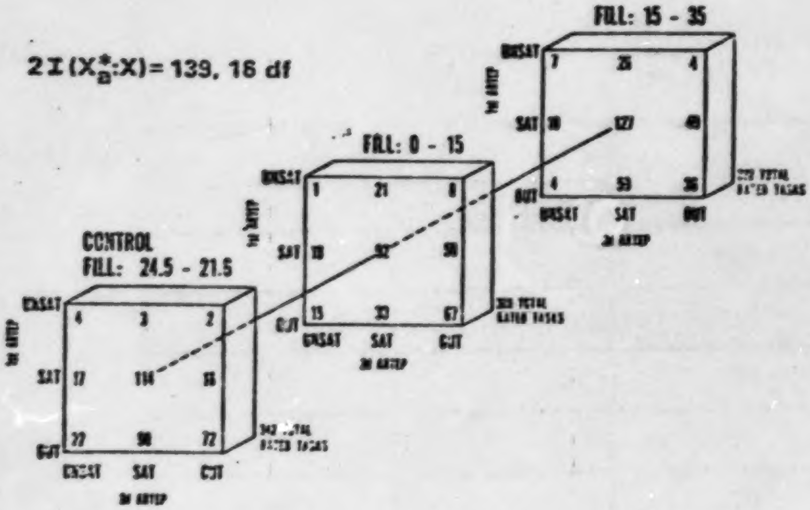


Figure B-6. Medical companies composite.

Together, the three medical units comprise a 3x3x3 contingency table. The question of interest concerns consistency of ARTEP scores across the 3 medical companies. More specifically, are the cell entries in the latter two medical companies consistent with those found in the control group? Since $2I(x_a^*:x) = 139.31$ with 16 df, we conclude that there is little consistency between the control group and the last two companies.

The Chi-Square critical value at $\alpha = .05$ for 16 df, is 26.296. Since this value is greatly exceeded in this case, it represents a high degree of dissimilarity between the control group and the other two units. The fact that the table sample sizes are different should be of no concern in arriving at this conclusion. This fact is taken into consideration when calculating the estimated cell frequencies under the null hypothesis of no difference. It should also be pointed out that the task items are not necessarily the same ones in the three tables, although they are nearly so. The fact that the table totals are different indicates that some tasks were excluded (not rated in both tests) or were not common to all three tables. On this account the premise must be made that all items are equally important for this kind of analysis to be of value. But the principal fact remains that the tasks were sufficiently alike to warrant such a comparison.

(2) With one exception, Figures B-7, B-8, B-9, and B-10 provide similar conclusions for the other four types of military units.

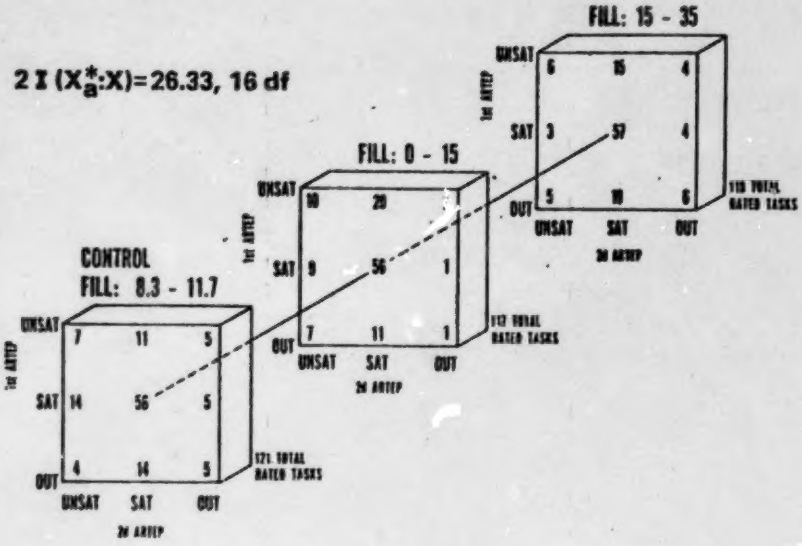

2 I (X*a:X)=26.33, 16 df
FILL: 15 - 35
UNSAT
SAT
OUT
1st ARTEP
6
15
4
3
57
4
5
10
6
UNSAT
SAT
OUT
2d ARTEP
118 TOTAL RATED TASKS
FILL: 0 - 15
UNSAT
SAT
OUT
1st ARTEP
10
20
9
56
1
7
11
1
UNSAT
SAT
OUT
2d ARTEP
117 TOTAL RATED TASKS
CONTROL
FILL: 8.3 - 11.7
UNSAT
SAT
OUT
1st ARTEP
7
11
5
14
56
5
4
14
5
UNSAT
SAT
OUT
2d ARTEP
121 TOTAL RATED TASKS

Figure B-7. Military Police companies composite.

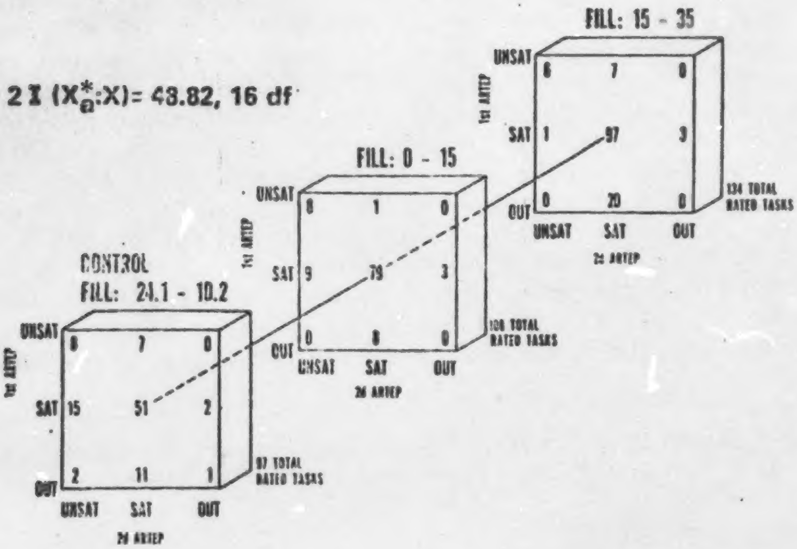

$2I(X_a^*:X) = 43.82$, 16 df
FILL: 15 - 35
UNSAT
SAT
OUT
1st ARTEP
6
7
0
1
97
3
0
20
0
134 TOTAL RATED TASKS
UNSAT
SAT
OUT
2d ARTEP
FILL: 0 - 15
UNSAT
SAT
OUT
1st ARTEP
8
1
0
9
79
3
0
8
0
108 TOTAL RATED TASKS
UNSAT
SAT
OUT
2d ARTEP
CONTROL
FILL: 24.1 - 10.2
UNSAT
SAT
OUT
1st ARTEP
8
7
0
15
51
2
2
11
1
97 TOTAL RATED TASKS
UNSAT
SAT
OUT
2d ARTEP

Figure B-8. Signal companies composite.

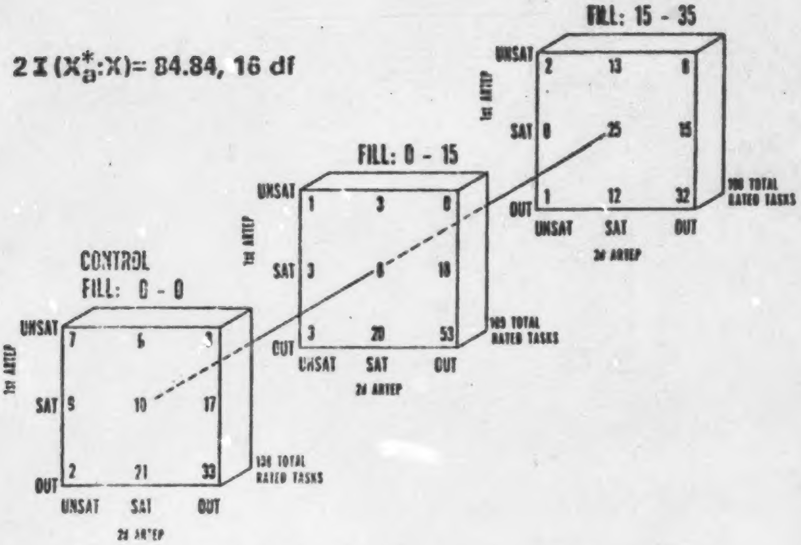

$2I(X_a^*:X) = 84.84$, 16 df
FILL: 15 - 35
UNSAT
SAT
OUT
1st ARTEP
2
13
8
0
25
15
1
12
32
108 TOTAL RATED TASKS
UNSAT
SAT
OUT
2d ARTEP
FILL: 0 - 15
UNSAT
SAT
OUT
1st ARTEP
1
3
0
3
8
18
3
20
53
109 TOTAL RATED TASKS
UNSAT
SAT
OUT
2d ARTEP
CONTROL
FILL: 0 - 0
UNSAT
SAT
OUT
1st ARTEP
7
6
3
9
10
17
2
21
33
108 TOTAL RATED TASKS
UNSAT
SAT
OUT
2d ARTEP

Figure B-9. Transportation companies composite.

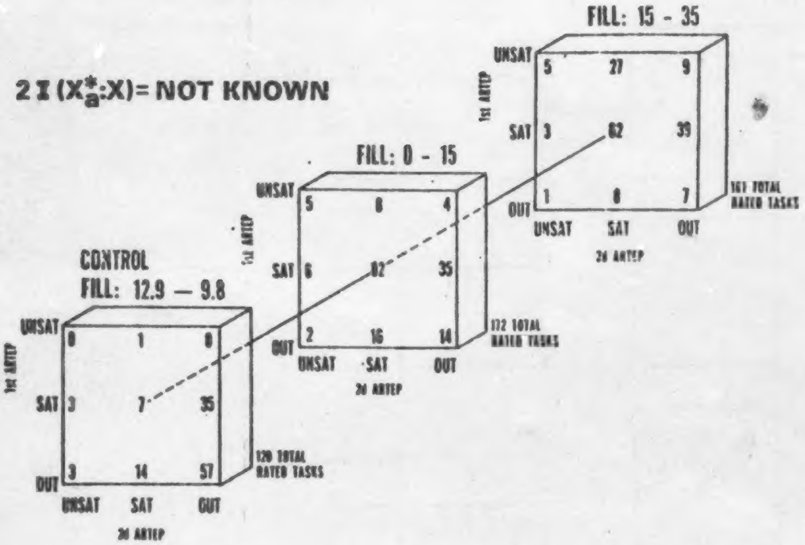

$2I(X_a^*:X)$= NOT KNOWN
FILL: 15 - 35
UNSAT 5 27 9
SAT 3 82 39
OUT 1 8 7
UNSAT SAT OUT
1st ARTEP
2d ARTEP
161 TOTAL RATED TASKS
FILL: 0 - 15
UNSAT 5 8 4
SAT 6 82 35
OUT 2 16 14
UNSAT SAT OUT
1st ARTEP
2d ARTEP
172 TOTAL RATED TASKS
CONTROL
FILL: 12.9 — 9.8
UNSAT 0 1 0
SAT 3 7 35
OUT 3 14 57
UNSAT SAT OUT
1st ARTEP
2d ARTEP
120 TOTAL RATED TASKS

Figure B-10. Maintenance companies composite.

The matrix for the maintenance unit could not be inverted so its MDIS was not obtained. However, an examination of the tables confirms that the ratings in each category are heterogeneous. A review of the $2I(x^*:x)$ values supports a finding that "the rating alignments within type groups are not homogeneous (consistent) with respect to the control group and are statistically different therefrom." The $2I(x^*:x)$ values for the 3x3x3 tables indicate that the changes in ARTEP scores, across ARTEPs, did not change in the same fashion for the control group as they did for the other two companies. That is, even within the same type of unit, the ARTEP scores fluctuated widely.

e. Control Groups.

(1) The analysis thus far has indicated great variation within ARTEP scores. This variation can be correctly described as "noise." A strong signal indicating the influence (either positive or negative) of female strength on unit performance has not yet been detected. To pursue this notion further, an examination was conducted of the stability of the control groups to assess whether the rating alignments of the control groups were stable between ARTEPs.

(2) Figure B-11 contains the test data for the 5 types of control groups.

$2I(X_a^{\star}:X)$

	1 \ 2	UN	SAT	OUT
TC	UNSAT	7	6	3
	SAT	9	10	17
	OUT	2	21	33

FILL: 0 - 0

−	0	+
29.63	46.30	24.07

0.35 2 df

	1 \ 2			
MP	UNSAT	7	11	5
	SAT	14	56	5
	OUT	4	14	5

FILL: 8.3 - 11.7

−	0	+
26.45	56.20	17.36

2.36 2 df

	1 \ 2			
MAINT	UNSAT	0	1	0
	SAT	3	7	35
	OUT	3	14	57

FILL: 12.9 - 9.8

−	0	+
16.67	53.33	30.00

12.72 2 df

	1 \ 2			
SIG	UNSAT	0	7	0
	SAT	15	51	2
	OUT	2	11	1

FILL: 24.1 - 10.2

−	0	+
28.86	61.86	9.28

13.60 2 df

	1 \ 2			
MED	UNSAT	4	3	2
	SAT	17	114	18
	OUT	22	90	72

FILL: 24.5 - 21.6

−	0	+
37.72	55.56	6.72

92.46 2df

Figure B-11. Summary of control units.

Since the level of fill was to be held constant for the two ARTEPs, the task ratings should reflect little or no change. Figure B-4 shows that 3 of the 5 companies were statistically different. For the maintenance unit, a net increase in performance was noted. However, for the signal unit a decrease was noted, and for the medical unit, as previously pointed out, the results were very unstable and performance decreased considerably in the second ARTEP. The ratings attained by the control groups are not stable between the two ARTEPs. In 3 out of 5 cases serious departures were noted. This finding cast doubt upon the utility of the ARTEP, as administered, as a suitable instrument for satisfying the primary MAX WAC test objectives.

f. Single ARTEP Units.

(1) There were 25 companies which received only one ARTEP. Figure B-12 aggregates the scores for these 25 companies and shows the percent of the total ratings by type unit and by category (outstanding, satisfactory, and unsatisfactory.)

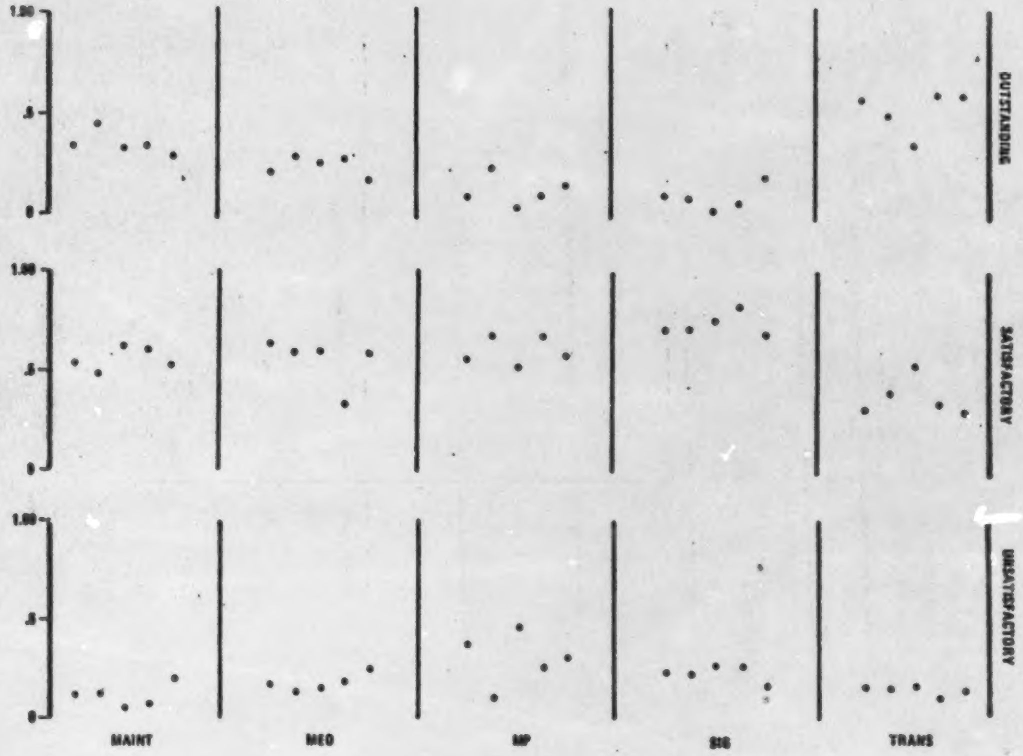


Figure B-12. Aggregate scores by type unit and category.

Since each unit was tested only one time, a cross-classification type of analysis could not be used. There are two inferences to be drawn from Figure B-12. First, ARTEP ratings vary greatly according to the type of company undergoing test. For example, consider the outstanding category. The five TC units scored a relatively high percentage of outstanding ratings while MP units received a much lower proportion of outstanding scores. Second, the percentages vary across the 3 categories, with the great majority (30 to 70% of the ratings, depending upon type units) being satisfactory. This type of variation in ARTEP scores makes it very difficult to detect small shifts in the scores due to female fill, should such shifts, in fact, exist.

(2) Figure B-13 depicts the same information, collapsed across type unit.

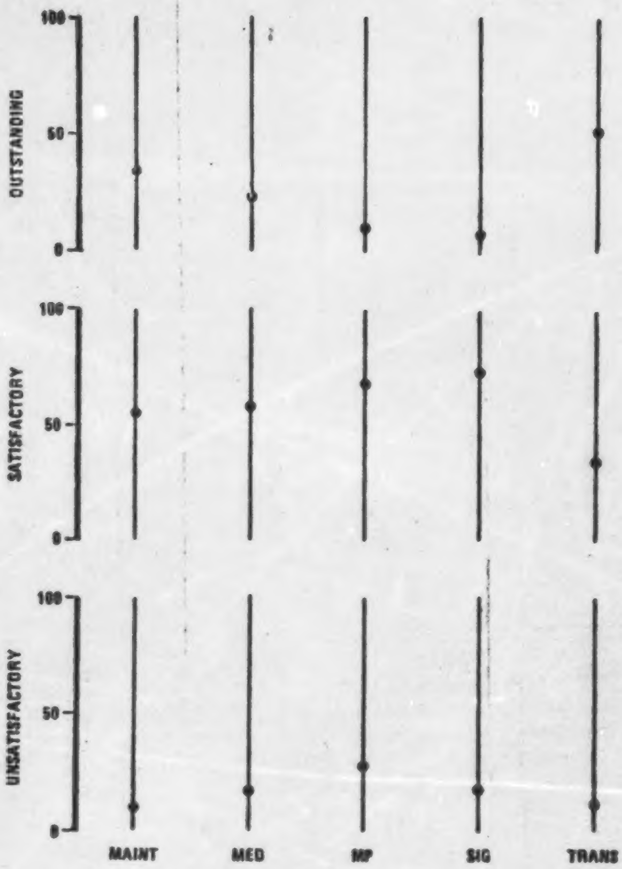


Figure B-13. Aggregate scores by type unit and category (single ARTEP units).

Again, the variation by type unit and rating category is easily observed. To test this notion the single ARTEP scores were cast into a 5x2x3 contingency table indexed as shown in Figure B-14. Results are shown in the Analysis of Information, Figure B-15, and are graphically displayed in Figure B-16.

FACTO[illegible]	INDEX	INDICES				
		1	2	3	4	5
TYPE COMPANY	i	MAIN	MED	MP	TRANS	SIG
FILL	j	LOW (< 10%)	HIGH (> 10%)			
RESPONSE	k	UNSAT	SAT	OUTSTAND		

Figure B-14. Contingency table index, single ARTEP units.

COMPONENT DUE TO	INFORMATION	DEGREES OF FREEDOM	CRITICAL LEVEL
A. X(ij.), X(..k) HOMOGENEITY	$2I(X:X_a^*) = 662.480$	18	$< 10^{-6}$
B. X(ij.), X(i.k) TYPE COMPANY EFFE[illegible]	$2I(X_b^*:X_a^*) = 587.936$	8	$< 10^{-6}$
INTERACTION	$2I(X:X_b^*) = 74.473$	10	$< 10^{-5}$
C. X(ij.), X(i.k), X(.jk) LEVEL OF FILL EFFECT/TYPE COMPANY	$2I(X_c^*:X_d^*) = 19.005$	2	$< 10^{-4}$
INTERACTION	$2I(X:X_c^*) = 55.460$	8	$< 10^{-4}$

Figure B-15. Analysis of information, single ARTEP units.

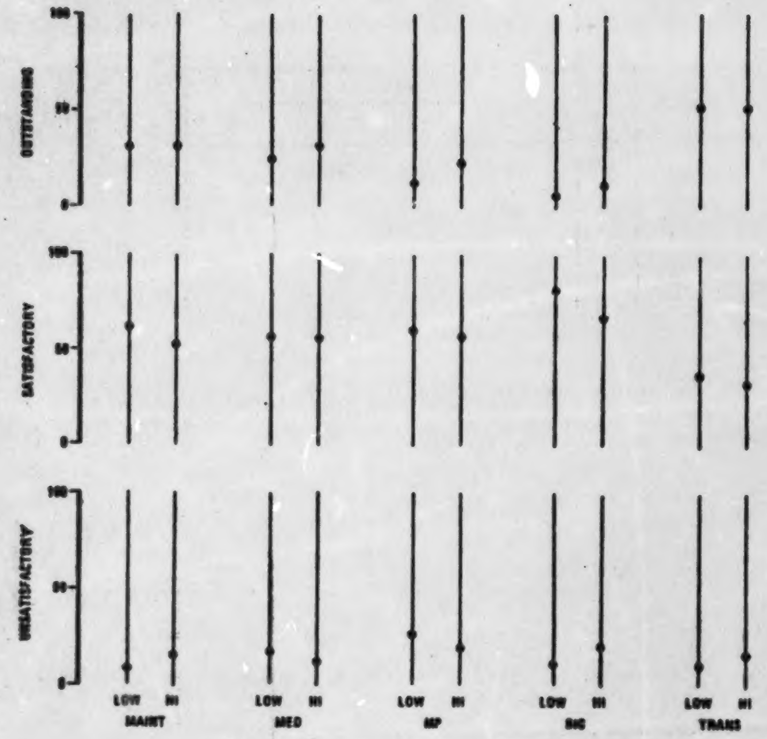


Figure B-16. Percent of ratings by category.

(3) Based upon 7,140 ratings summarized in Figure B-17, several questions can be asked of the data.

TYPE COMPANY			MAINT		MED		MP		SIG		TRANS	
LEVEL OF FILL			LOW	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH
RESPONSE	OUTSTANDING	3	199	206	181	690	119	32	17	82	283	171
	SATISFACTORY	2	396	344	415	1243	576	78	220	456	202	106
	UNSATISFACTORY	1	54	103	126	283	257	26	30	137	57	51

Figure B-17. Summary of ratings, single ARTEP units.

The most important question is, "Are the percent of unsatisfactories, satisfactories, and outstanding ratings awarded relatively uniform or consistent across type of unit and level of fill?" This question can be addressed in the Analysis of Information Table, Figure B-15, which is similar to an analysis of variance table. The null hypothesis of homogeneity is easily rejected since $2I(x:x_a^*) = 6,624.08$ with 18 degrees of freedom, is highly statistically significant. This indicates that either type of military unit, or level of fill, or perhaps both, may be affecting the response variable (percent of ratings by category). Examined in the light of this statistical evidence, a finding that the type of unit and level of fill do influence the percent of rating by category may be possible; however these findings must be further tempered by the injunction raised earlier concerning the impact of other extraneous factors upon the data. The impact of these extraneous variables could have caused perturbations in the data which were detected by the statistical analysis.

(4) The division between low and high fill seen in Figure B-17 (less than 10% and greater than 10% was arbitrary and may have influenced the outcome) indicates that units with the greater percent of females appear to perform better than those with less. This difference, although statistically significant, is very small as shown in Figure B-18.

	LOW FILL		HIGH FILL	
RATING CATEGORY	NUMBER OF RATINGS	PERCENT	NUMBER OF RATINGS	PERCENT
OUTSTANDING 3	799	25.5	1181	29.5
SATISFACTORY 2	1809	57.8	2227	55.6
UNSATISFACTORY 1	524	16.7	600	15.0
	TOTAL 3132		TOTAL 4008	

Figure B-18. Percent change in high-low fill by rating category.

However, the main point to note is the great variation between military type units. The relative magnitude of the "unit effect" is roughly 30 times that of the "fill effect." This suggests that the type unit is a far more important consideration than the level of fill, at least for those kinds of units and levels encountered in this analysis.

g. The primary conclusion to be drawn from this statistical analysis is: The noisy data, great variation in ARTEP scores within types of tested units, and the instability of the control groups, strongly suggest the presence of extraneous variables which could not be controlled statistically and which were not controlled during the administration of the test. This conclusion cast serious doubt upon the utility of the ARTEP, as administered, as a suitable instrument for satisfying the primary MAX WAC test objectives.

TAB C

QUALITATIVE ANALYSIS

1. Discussion. As a part of the OTEA visits to units which participated in MAX WAC ARTEPs, an independent judgmental assessment of subjective factors affecting MAX WAC, was made by military members of the team. The following methods were applied to this assessment:

a. Unstructured discussions were held with personnel who had participated as players, local command evaluators or controllers. Details of the results are summarized in paragraph 2a below.

b. After action reports prepared by the chief evaluators after each test were reviewed to identify factors or conditions in the test which the evaluator considered unusual, and which could have affected test data. Results are summarized in paragraph 2b below.

c. Although not part of the assigned purpose, the team nevertheless gained considerable insight into the perceptions of the MAX WAC participants concerning the advantages and disadvantages of having female soldiers assigned in significant numbers. These are summarized in paragraph 2c below.

2. Analysis of Observations.

a. In visiting the five units (which accounted for a total of seven tests), the following factors and conditions were found to have varied from normal or controlled levels to an extent that an effect on ARTEP performance appeared likely.

(1) In all units, the NCO structure was predominantly or entirely male. One unit had 3 female Sp 5's, none in supervisory roles. Another had 2 female acting sergeants. Other than that, all other enlisted women in the units visited, appeared to have been grade E-4 and below. This is a natural consequence of the recent entry of women into most of these MOS's and type units however, it is considered unrepresentative of the steady state condition that will exist when women have advanced in normal career progression. Its effect on test results lies in the inexperience of male NCO's in directing women (another factor that can be expected to correct itself with time). The team observed in the field and perceived in discussion, that the male NCO's tended to let the women get by with minor acts and omissions that they would not permit their male soldiers, partly from lesser expectations and partly from shyness or misplaced gallantry. They also tended to assign tasks first to men and not really attempt to use women until the men were fully committed. Thus, it is reasonable to suspect that female soldiers were not fully utilized in the ARTEP, as compared with their potential utilization.

(2) The workloads were not consistent between units or tests. In two of the units (transportation and military police) it was generally felt that the scenario had taxed them to the limit. The medical unit leaders stated that the scenario exercised their full capability, but it did not appear to the observer team that individual unit personnel felt the work load had pushed them to the limit of their ability or endurance. The maintenance unit did not appear to be pushed to its full capacity, (approximately 30% utilization) primarily due to the difficulty in finding enough representative items for maintenance/repair work. The effect of the latter two instances, combined with the second priority use of the women noted elsewhere, is to create a perception which tends to minimize the contribution of women to the unit's ARTEP performance. The most serious work load effect observed, however, was that of a signal company (which took two ARTEPs). In the first instance, this unit was tested in the course of a division CPX and was under pressure to satisfy actual communications requirements under the direct scrutiny of the division commander. The second test was taken in isolation, with a command attitude that the test was only to satisfy MAX WAC requirements. The performance requirements and motivation were therefore drastically different.

(3) The extent to which different individuals' and units' normal garrison activities contributed to or detracted from their readiness for an ARTEP differed widely. The signal and medical units visited were divisional units which regularly went to the field in support of the division. The units and their personnel were, therefore, fairly regularly exercised in essentially the same activities as tested in the ARTEP. The other three type units were nondivisional units which normally performed garrison support missions that were markedly different from the ARTEP tasks. These units went to the field far less frequently. It was observed that in a unit trained primarily for garrison maintenance, tasks such as setting up a maintenance tent, were tasks assigned only to the men. However, once the unfamiliar phase was over and a task, such as the maintenance job normally done in garrison, was started, the women again become effective members of the organization.

(4) The NCO's of all units appeared also to be less certain in their dealings with women than with men. Under circumstances where the unit tasks were somewhat unfamiliar, coping with both the newness of tasks as well as the presence of women further reduced effective utilization of women.

(5) In both instances when double ARTEP companies were visited, it was found that for the second ARTEP the local command had made extensive last minute efforts to fill the companies to a higher level of female soldiers at the expense of the continuity of normal working or personal relationships. In both instances, a number of women had been placed in the unit as little as three weeks before the ARTEP, some by attachment only until completion of the ARTEP. Many of these women, while working in their primary MOS, came from jobs in which they had not been using that MOS or had been performing their MOS duties in a different manner or on

different equipment. Many women were directly out of AIT. In most cases, they displaced men who had been doing the job and whose aptitudes and limitations were known to their supervisors. In the one ARTEP, the effect of this lack of continuity was so evident that it was generally not even necessary to ask which women were newly assigned or attached; they were the ones who were being ignored. Since the first ARTEP in both instances was taken with personnel who had come to the unit through normal assignment procedures, it is considered that the artificial assignment procedures used in the subsequent ARTEP tended to negate a valid comparison between the two ARTEP's.

(6) Two units showed evidence of poor leadership. This was manifest by an apparent failure to recognize or deal with complaints relating to normal hardships that are inherent to the combat situation which the ARTEP seeks to reproduce. These complaints were made by both men and women. In both cases, a change of command occurred shortly after the ARTEP. One of these was a specific relief for cause and, while not clear in the other case, it is the opinion of the senior officer of the observer team that an attitude problem of sufficient magnitude had existed in the unit at the time of the ARTEP which would have made the change of command necessary. Both units in which a leadership problem was identified were single ARTEP units. The lower scores in these two units, as compared with other units, might be used to draw inferences about the effect of their content of female soldiers, when in fact the quality of leadership was probably the dominant factor.

(7) In one case it was evident that some dissension had existed between the local command's controllers/evaluators and the MAX WAC evaluators. The local command felt that they were the ones who had been tasked to execute the scenario and that the MAX WAC people came late on the scene with detailed interference and lack of coordination. While the test was evidently executed satisfactorily, this friction was visible to the test unit, affecting their attitude and exposing them to some additional harrassment. Examples were directing a female soldier to change a truck tire as a separate exercise, even though there was said to be ample opportunity to observe this in the course of test events, and conducting a second NBC attack because the MAX WAC evaluators had not been in position to observe the first one. Since only the local command's side of this was heard, no attempt was made to assess the accuracy of these complaints, or determine fault. However, it should be noted that the observed friction and lack of coordination evidently did have a negative effect on the unit's attitude.

(8) One unit with a requirement for a 35% female fill, had only two female NCO's in a relatively high grade enlisted rank structure. The effect of meeting the MAX WAC test design fill requirement for the second ARTEP, was to fill the lower 4 grades to nearly 60% with females. This, combined with the previously noted condition of the recent assignment of many of the women, introduced a further degrading factor as compared with the unit's first ARTEP. The artificial effect of such a high percentage of women in the lower grade structure cannot be used as an indicator of the results which could be obtained with a more uniform fill made over a longer period of time.

b. In assessing the extent to which the MAX WAC test met its specified objectives, fifty-five ARTEP narratives written by team chief evaluators from the MAX WAC Directorate, were examined. Team chief evaluators commented on factors they considered significant during the conduct of the ARTEPs. The following is an analysis of the factors which could influence ARTEP results. A summary of the number of tests in which the evaluator felt that a situation existed that was sufficiently aberrant as to merit comment, is shown in Table C-1.

TABLE C-1. VARIABLES AFFECTING ARTEP SCORES

TYPE UNIT	NATURAL FACTORS	LEADERSHIP & ORGANIZATION	WORK LOAD	HIGHER COMMAND POLICY	TRAINING & FIELD VS GARRISON MISSION	PERSONNEL STABILITY	ASSIGNED/PFD ARTEP STRENGTH	SITE EVALUATION EFFECTIVENESS	EQUIPMENT SHORTAGES/ OPERATIONAL	MOS SKILLS	PRE-ARTEP SCENARIO PRACTICE
MAINTENANCE	1 –	1 + 5 –	1 + 5 –	5 –	3 –	1 –	4 –	3 + 2 –	1 –		
MEDICAL	1 –	2 + 2 –		4 –	3 + 3 –	1 –		2 –	2 –	1 –	2 –
MILITARY POLICE	4 –	4 –	1 –	1 –	1 + 2 –		1 –	6 –			
TRANSPORTATION		4 –		4 + 2 –	5 –			1 –		1 –	
SIGNAL	4 –	3 + 4 –	2 –	2 –		1 +	3 –	2 –	1 –	1 –	

" + " INDICATES A POSITIVE OR FAVORABLE INFLUENCE
" – " INDICATES A NEGATIVE OR ADVERSE INFLUENCE

(1) The single most important factor is considered to be quality of leadership and effective organization. Units with experienced company commanders who demonstrated outstanding leadership ability, generally performed better than units with weak leadership and/or poor organization. Strong leadership on the part of platoon leaders, first sergeants, and platoon sergeants, is also a major factor in the success of a unit. For example, in one unit both the battalion commander and the company commander had been in command for a short time. As a result, both were apprehensive about undergoing an ARTEP observed by a DA Team and demonstrated somewhat less-than-dynamic leadership during the ARTEP.

(2) Higher command policy is considered to be a dominant factor affecting ARTEP performance. At installations where the command structure had a positive attitude toward utilization of female soldiers, the attitude permeated down through command levels. This created an atmosphere wherein female soldiers were treated like mature adults and given an opportunity to work in their MOS. Problems were anticipated and

resolved as they arose. In other cases, some installations heavily tasked MAX WAC units with garrison support missions without regard to their upcoming ARTEPs. This greatly impaired unit preparation, particularly in those units where a higher female fill required time to assimilate.

(3) Adequacy of field training in the months prior to ARTEP varied considerably from unit to unit. There were several units where no field training had been conducted in almost a year and other instances where a particular section had not been to the field in several years. In one unit, the supply sections had not been to the field in over a year and consequently had poor scores in warehousing tasks. There were also instances where maintenance, medical, MP and signal units normally performed garrison missions which were considerably different from the field (ARTEP) mission. For example, there was a General Repair Section with a field mission of repairing power generator equipment. However, because of other diverse garrison maintenance assignments, this unit had not performed power generator maintenance on a regular basis.

(4) MAX WAC ARTEPs evaluations were carried out by local evaluators provided by next higher headquarters. The effectiveness of the evaluation varied depending on attitude of evaluators, relative experience of evaluators, cooperation between local evaluators and MAX WAC Directorate evaluators, and adherance to scenario sequence. There were instances where evaluators demonstrated a very negative attitude toward the MAX WAC test, eliminated important tasks from the scenario, did not cooperate with MAX WAC Directorate evaluators. There was one case where no operations order was given to the unit.

(5) Where factors such as adequacy of training area or weather conditions were substantially different, these adversely affected the comparison of ARTEP performances. One training area consisted of a single hardtop road and only a few unimproved single lane roads. This was only marginally suitable for a training exercise requiring tactical road marches, area patrol and land navigation. In cases of weather related factors, winds in excess of 35 knots made it difficult, in one instance, to erect antennas. Extremely cold weather, low wind chill factor, and heavy rains caused severe problems in several other cases.

(6) Sufficient workload is a necessary element of an ARTEP. Insufficient workload did occur in many cases. In a General Repair Section, for example, no repair work was observed to be taking place. In addition to the workload, equipment shortages existed. One electronics maintenance section was short test equipment and could not be evaluated. One medical unit's X-ray equipment was inoperative.

(7) It must also be noted that quality of MOS training affected the ARTEP. Self-paced AIT courses enabled some women to complete AIT sooner than normal. This can cause problems as it did with one TC unit where an MOS 64C vehicle mechanic had never learned to change a 2 1/2 ton truck tire during AIT.

(8) Although most units met their 80% ± 10% personnel strength requirement, some units were well understrength. Degradation of platoon and/or section strength was detrimental in a few instances. Stabilization of filler personnel 60 days prior to the ARTEP was rarely accomplished in the control groups. Several units received personnel only a few days prior to ARTEP.

(9) Although the content of ARTEPs are known by all units, the scenario for each specific test is not. However it was learned that two units obtained the scenario for their special ARTEP and practiced it prior to the actual test. It can be assumed that these units were better prepared for the ARTEP and obtained scores of questionable value to themselves as well as MAX WAC.

c. In the course of visiting the five units described in para 2a above, the team discussed with the personnel of the units a number of their perceptions concerning the advantages and disadvantages of females in the unit. While these are not germane to the validity or comparability of the ARTEPs, they are relevant to the questions which the MAX WAC test seeks to address. Note that during discussion with female soldiers at field sites, lack of durability of the fatigue uniform was mentioned as a persistent problem. In TOE units, where fatigues are the duty uniform, life expectancy is much less than the more durable fabric in male fatigues. In fact, most female fatigues were cited as lasting only 7 months. The problem is further complicated with the realization that there is no female wash and wear fatigue uniform and the fatigues presently available must be starched to look good.

(1) Perceived advantages of women in the unit.

(a) All five units indicated that women generally performed better than men in some tasks. These were generally tasks involving attention to detail. The Military Police unit indicated that women were essential for some tasks and that, in fact, before female MPs were available they had had to borrow the services of other women, such as nurses, to assist. It was also indicated that women were more effective in some interview situations.

(b) All five units indicated that women were less likely to be disciplinary problems. They did not tend to get into minor troubles caused by such factors as excessive drinking or fisticuffs. One commander observed that when they did get into trouble, it would be something more serious, but there was no indication that serious trouble would be more frequent than with the men.

(2) Perceived disadvantages of women in the unit.

(a) The most strongly expressed concern by the commanders of all units visited was the loss of time and deployability due to pregnancy. Estimates, not supported by data, were that if a unit had over about 30%

women, loss due to pregnancy would significantly degrade their operational readiness. It was also stated (again, without supporting data) that about half the pregnancies were with unmarried women, yet in no case was there any evidence of a command effort to discourage or help prevent or terminate these unmarried pregnancies. The team did not find any evidence of policy guidance at the unit level as to what a commander could do in the way of advice, moral suasion or medical assistance. In the absence of any such guidance, commanders were understandably reluctant to touch the subject, even though they identified it as their most serious concern with female soldiers.

(b) All units visited identified male NCO leadership as a problem area. As noted in para 2a(1), almost all of the females were grade E-4 or below and almost all of the NCO structure was male. The male NCOs for the most part, were less effective in dealing with their female soldiers than their male soldiers, expecting and therefore getting less performance from them and allowing them to get away with things that they would not permit their male soldiers to get away with. The extent to which this may have been true of male junior officers was not observed for several reasons. One is that they were second or third line supervisors of most of the women so had less direct contact. There was also a female officer in each of the companies visited and there was a subconscious (in one case, conscious) tendency to shift the burden of uniquely female leadership problems to her, regardless of whose responsibility the problem soldier might actually be. In all the units visited, only one male officer reported having ever had any specific instruction in female leadership, which he said was most valuable to him.

(c) In four of the five units visited, commanders perceived that women would be less able to endure prolonged stress than men. This was not supported by systematically gathered data, but cases of exercises in which some of the women had in fact been less durable, were cited. The perception also appeared to be based on the womens' greater concern for cleanliness, privacy, and need for sanitation. It was also acknowledged that the weakness in male leadership previously noted may have resulted in a lower level of motivation of the women, compared with the men.

(d) In four of the five units visited, there was general agreement that the strength requirements of some tasks exceeded the strength of many of the women. Examples were handling the lifting tackle of a recovery vehicle, carrying litters, changing large truck tires and setting up large antennas. The usual solution was to allocate enough men to the various sections to insure that men were available for those tasks or to use two women where one man might have sufficed. In some cases the NCOs had to perform some of the womens' tasks. It was also noted that some of the jobs or equipment could be re-engineered to reduce the strength requirement.

(e) Four of the five units complained that the women were less well trained in the non-MOS soldierly skills. There was a general perception that the male basic training had been more demanding and more comprehensive than that of the females. The women had little knowledge of individual or small unit combat techniques or of crew served weapons and, particularly in the nondivisional units, there had been little opportunity or effort to provide that training. The weakness in male NCO leadership also operated against improvement in this area.

TAB D

FOLLOW-ON EVALUATION OF LONG TERM STRESS SITUATION

1. Discussion.

a. To evaluate the relative performance of male and female soldiers under conditions of extended stress, a team of OTEA personnel visited a selected long term, free play exercise as a follow-on to the analysis of data collected in the MAX WAC evaluation. The purpose of this visit was to observe female performance in an extreme environmental condition as well as to evaluate their performance on an extended exercise.

b. The team, consisting of two male senior officers (O-6), a female officer (O-3) with successful field command experience, and two DAC, a male research psychologist and a female systems analyst, visited the Opposition Forces Logistic Support Activity (LSA) and Joint Headquarters (JOPFOR) areas of Exercise BRAVESHIELD at USMC Base, Twenty Nine Palms, CA 17-18 July 1977. Except for selected senior personnel involved in the test, the team visit was not made known in advance. The team visited the units listed in paragraph 2c below. Discussions were initially held with officers (usually O-3 or lower) and then team members circulated as individuals or in groups of two or three, talking with male and female soldiers at their work sites or in their tents. After the team member had stated the purpose of the visit, troops were encouraged to discuss in a totally unstructured manner, their life style during the exercise, relationship with their peers, supervisors, or subordinates, particularly of the opposite sex, job requirements and performance, problems, annoyances, etc. Personal interactions, job performance, and life styles of the soldiers were observed. Impressions and information acquired by the various team members were discussed among themselves and, as appropriate, follow-up visits and observations were made. Observations were over a period of two days and discussions were held with between 100 and 150 people, about half of whom were women and most of whom were in the lower enlisted ranks.

2. Results.

a. Exercise Environment. The area visited is an extremely remote one in the NE portion (area Echo) of the USMC Base, Twenty Nine Palms, CA. It is entirely void of any facilities, either military or civilian. Topography is rocky desert plains and lava outcroppings rising to jagged barren mountains. The sparse vegetation consists of widely scattered cactus and weeds with nothing over two feet high. Daytime temperatures were consistently in excess of 100°F, usually over 110° and frequently over 120°, falling into the 80°'s at night. High afternoon winds (thunderstorms and sandstorms) brought little temperature relief but many emergency tent repairs.

b. Living conditions.

(1) The LSA was setup nontactically to support the exercise opposition force. General purpose tentage, from pyramidals through G.P. large, was used for most living and working areas. Limited electrical power from motor generators and field lighting sets was available. Water was readily available (lukewarm) from lister bags throughout the area. The mess halls were supplied with ice and a limited amount (enough for about one picnic cooler per 5 or 10 person section) was made available to the troops. Incident to required trips into the base, most sections were able to maintain a limited supply of soft drinks. There was little beer and no evidence of any hard liquor. There were no mobile PX services or field clubs. Mess halls served a "B" ration for breakfast and "C" rations were issued to individuals for all other meals. A shower point was established in the area, with blocks of time set aside for use by women. Some sections also had individual gravity shower units. The engineers had dug pits and provided outhouses for latrines but these were inadequate in number and capacity and difficult to keep deodorized.

(2) The JOPFOR Hq was about three miles from the LSA and was set up tactically with facilities dispersed and well camouflaged. Principal elements were a TOC (serving an 07 OPFOR commander), MI elements, a DASC and an extensive communications complex. Only the minimum essential people were billeted in this area, with most commuting from the LSA.

c. Units visited (most units and personnel had been on site since 1 July, all since 9 July).

(1) 9th Signal Battalion (Ft Lewis) had approximately 200 personnel in the field of whom about 40 were female, the senior being a 1st Lieutenant. The provisional organization was formed by augmentation to the battalion's B Company and its mission was to provide division level communications to the JOPFOR under direction of the battalion S-3.

(2) Provisional Detachment, 11th Signal Group (Ft Huachuca). This appeared to be entirely provisional in nature, operating under the direction of the JOPFOR, J6, to provide Corps and joint communications to the JOPFOR. It had approximately 100 personnel of whom about 12 were female, the senior being a 2d Lieutenant.

(3) The provisional military intelligence detachment was a mixture of regular and USAR elements from diverse locations and its personnel included both regular and reserve female soldiers.

(4) A Co, 7th Medical Battalion (Ft Ord) was the only unit visited that was operating in its TOE configuration. It had about 15 percent women up to grade E-6. Two doctors were attached and the unit was charged with medical support of the JOPFOR.

(5) HHC, 1st Bde, 9th Inf Div (rear) (Ft Lewis). Elements of this unit, located in the LSA, were heavily augmented to provide DISCOM type services to the JOPFOR. This included attachment of about 12 women (senior being a 1st Lieutenant) to this previously all male organization. Senior officers with whom discussions were held were the S-1 and the Chaplain.

d. Summary of discussions and observations.

(1) Peer acceptance of female soldiers. One of the most consistent and impressive findings was the acceptance of the female soldiers, as soldiers and as partners in their work and their life style, by their male peers. The men in the sections evaluated the women they worked with according to their ability, just as they did their male peers, and having women in the section was simply "no big deal." The extended period of shared hard work, deprivation and discomfort had done away with any feelings of strangeness or gallantry or any toleration of any member doing less than his or her share. The fact that in this environment the women had earned acceptance attests that women did adjust to the requirements of the situation to about the same extent as the men did, and that the women did perform up to their individual job requirements.

(2) Supervisory acceptance of female soldiers. Supervisory reactions paralleled that of peers to the extent that the women were regarded as having done generally as well as the men in those jobs to which the women were assigned. This was qualified by the fact that in job assignments, the supervisors had given consideration to what they considered to be the strength limitations of the women; e.g., women were assigned as radio operators but not as cable layers. No commander expressed any concern about being unable to accomplish his mission due to female soldiers. There had also been problems as to privacy and personal hygiene (see paragraph (3) below) some of which, it was generally conceded, could have been avoided if they had been anticipated.

(3) Female acceptance of the exercise situation. It was apparent from all categories of comment (supervisors, male peers and female) that the severity of the situation came as more of a shock to the women than to the men. They had gone into the exercise with less of an idea as to what the exigencies of the situation would be or knowledge of ways to cope with the situation. Adjustment to these stresses seemed to have taken a few days longer than for the men, because of the failure of commanders to properly indoctrinate them, but was completed by the time of the OTEA team visit. The women indicated that they accepted and could cope indefinitely with the situation. It was noted that most of the women continued to keep themselves well groomed, much more so than the men, some still wearing make up, washing, combing out and putting up their hair, using skin cream and so forth. This effort appeared to be appreciated rather than resented by their male peers and may also have positively influenced male hygiene. A significant female complaint that remained at the time of the visit concerned privacy. Some of this was

due to the required proximity of tent living and some due to the restrictions required to gain privacy. Most of the units had provided a separate tent for the women but, due to the weather, discomfort was severe if the sides were not rolled up. Rolling up the sides of the tents minimized privacy for both females and males alike. After the first few days many of the women elected to billet with their duty sections, that being more convenient and there being little difference in privacy. (This was standard practice in the Medical Company from the beginning.) The inadequate latrine situation required sharing of latrines, with need for latches, waiting in line, male escorts and other embarrassing and inconvenient conditions. The offensive condition of many of the latrines bothered the women more than the men. Some women complained of the difficulty of personal hygiene during the menstrual cycle. The problem of hygiene and menstrual discomforts could be greatly minimized by making better feminine hygiene products, analgesics, and packaged towelettes readily available.

(4) Physical and medical problems. The only uniquely female problem reported by medical personnel were some complaints of early, heavier menstrual flow and somewhat worse cramps, all of which were classified as due to the severe heat and none of which interfered with the duties. There was no significant difference reported in resistance to heat exhaustion, with men and women perceived as being affected approximately in proportion to their numbers. The rate for either was surprisingly low. There also did not appear to be any difference in the rate at which men and women had to be evacuated from the field for other than injuries.

(5) Social relationships. There was no evidence that the presence of women created any serious social problems. It was known that sexual intercourse was occurring, but not more than occurs in garrison. The heat, lack of privacy and wide open terrain were credited with reducing both the incentive and the opportunity. The team neither observed anything nor received any comments indicating that promiscuity was a problem. In the area of unwanted attentions, there had been a problem with vulgarity directed at the women and some prurient interest early in the exercise. Much of this had come from an infantry battalion bivouaced next to the LSA. It illustrates, that this type of problem can be expected when female soldiers have to deal with units that have no females or experience with females as soldiers. With the departure of the infantry battalion and the remaining males' acceptance, this was no longer considered a problem as the exercise continued. In fact, some commanders indicated that the men became protective of the women in their units regarding unwanted attentions from men in other units.

(b) Combat Expectations. The team was not able to observe the performance of non-MOS related combat tasks and there was no particular awareness among the combat service support troops that what they were participating in was intended to be a simulation of combat. Many of the troops, particularly the women, had not thought it through to realization that had it been a war, both male and female soldiers could have been killed or wounded, or that they could have killed or wounded enemy soldiers. Realization that this was the ultimate purpose of what they were doing appeared to come as a shock to some of the young female soldiers. Again, this is a lack of proper indoctrination by commanders.

3. Potential of long term, free play exercises for future evaluations.

a. Advantages.

(1) Allows stabilization of the supervisory and peer relations under the particular set of field conditions. Indications in this test were that this took from three to six days.

(2) Allows observations of both short term stress (by observing situations of intense activity in the early phases) and long term stress (by observing the later phases and periods of grueling, tedious activity).

(3) Presents a plausibly realistic profile of the required activities (assuming that a realistic scenario and exercise play are utilized), especially for combat service support units.

(4) Minimizes burden on troops. This assumes that advantage would be taken of already planned exercises and that no extra troop activity would be written into them for this evaluation.

(5) Does not require a large directorate in the field. In that, validity of results depends on spontaneous or natural responses of the soldiers, a large or highly visible establishment in the field could be self-defeating.

b. Disadvantages.

(1) Does not assure that all aspects of job performance are evaluated. A penalty of the realistic task profile is that the particular situation may not require all the skills of the MOS, or may not exercise some non-MOS skills.

(2) Most raw data will be subjective. Insuring objective results that can withstand critical review will require the greatest care and skill in selecting and training data gatherers and in data reduction and analysis.